

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Brenwen Loeb Examiner #: 78225 Date: 2/2/01
 Art Unit: 1636 Phone Number 301 605-1197 Serial Number: 09/000,002
 Mail Box and Bldg/Room Location: 411A-09 Results Format Preferred (circle): PAPER DISK E-MAIL
11E-12

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Nucleic acid fragments recombinant vector containing the same and method for
promoting the expression of structural gene by using the same
 Inventors (please provide full names): _____

→ UEKI, Jun

Earliest Priority Filing Date: 11/19/98

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Please search SEQ. ID. NOS. 1-3 in the
 nucleotide databases. Don't search pending
 patents. Thank you.

POINT OF CONTACT:
 BARB O'BRYEN
 TECH. INFORMATION SPECIALIST
 STIC CM1 12C14 308-4291

STAFF USE ONLY

Searcher: BOB
 Searcher Phone #: _____
 Searcher Location: _____
 Date Searcher Picked Up: 2-5-01
 Date Completed: 2-13-01
 Searcher Prep & Review Time: 15
 Clerical Prep Time: _____
 Online Time: 12

Type of Search

NA Sequence (#): 3
 AA Sequence (#): _____
 Structure (#): _____
 Bibliographic: _____
 Litigation: _____
 Fulltext: _____
 Patent Family: _____
 Other: _____

Vendors and cost where applicable

STN: _____
 Dialog: _____
 Questel/Orbit: _____
 Dr.Link: _____
 Lexis/Nexis: _____
 Sequence Systems: 15
 WWW/Internet: _____
 Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 20:28:44 : Search time 136.04 Seconds
(without alignments)
176,730 Million cell updates/sec

Title: US-09-600-602-1

Percent score: 64
1 taagccagtgcttaagc.....gcttctcttcgcgctca 64

Sequence: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Scoring table: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36:*

1: /SID56/gcgdata/geneseq/NA1980.DAT:*

2: /SID56/gcgdata/geneseq/NA1981.DAT:*

3: /SID56/gcgdata/geneseq/NA1982.DAT:*

4: /SID56/gcgdata/geneseq/NA1983.DAT:*

5: /SID56/gcgdata/geneseq/NA1984.DAT:*

6: /SID56/gcgdata/geneseq/NA1985.DAT:*

7: /SID56/gcgdata/geneseq/NA1986.DAT:*

8: /SID56/gcgdata/geneseq/NA1987.DAT:*

9: /SID56/gcgdata/geneseq/NA1988.DAT:*

10: /SID56/gcgdata/geneseq/NA1989.DAT:*

11: /SID56/gcgdata/geneseq/NA1990.DAT:*

12: /SID56/gcgdata/geneseq/NA1991.DAT:*

13: /SID56/gcgdata/geneseq/NA1992.DAT:*

14: /SID56/gcgdata/geneseq/NA1993.DAT:*

15: /SID56/gcgdata/geneseq/NA1994.DAT:*

16: /SID56/gcgdata/geneseq/NA1995.DAT:*

17: /SID56/gcgdata/geneseq/NA1996.DAT:*

18: /SID56/gcgdata/geneseq/NA1997.DAT:*

19: /SID56/gcgdata/geneseq/NA1998.DAT:*

20: /SID56/gcgdata/geneseq/NA1999.DAT:*

21: /SID56/gcgdata/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	64	100.0	173	17	Promoter #2 for ph
2	64	100.0	173	19	ubiquitin gene int
3	64	100.0	183	17	Promoter #1 for ph
4	64	100.0	2799	16	DNA encoding Phosp
5	64	100.0	2799	17	Phospholipase D ge
6	24.8	38.8	1331	8	DNA sequence encod
7	24.8	38.8	2110	19	Mouse Friend virus
8	23.6	36.9	1037	20	Human secreted pro
9	23.6	36.9	2306	11	Sequence encoding
10	23.6	36.9	2306	11	Murine IL-2R beta
11	23.6	36.9	7947	20	Enterococcus faeca
12	23.4	36.6	508	18	Hamster Ubiquitin/

13	22.8	35.6	219	20	Z20473	Barley microsatell
14	22.8	35.6	2353	19	V455981	A. thaliana sterol
15	22.8	35.6	2464	18	V03752	Porcine TNFalpha-c
16	22.8	35.6	5888	20	Z24013	Human GDNF fragmen
17	22.6	35.3	3155	17	T42120	M-lats2 gene encod
18	22.6	35.3	3155	21	Z51507	Mouse lats2 (large
19	22.4	35.0	3336	6	MS0149	Sequence of the cd
20	22.2	34.7	332	16	T19131	Human gene signatu
21	22.2	34.7	1738	19	V09299	Nucleotide sequenc
22	22.2	34.7	9558	16	O88328	Valencia orange rl
23	21.8	34.1	380	20	V86878	EST clone BK241.
24	21.8	34.1	1228	20	X82099	Human calcium-acti
25	21.8	34.1	1608	21	Z22258	Human potassium ch
26	21.8	34.1	2417	20	X91430	T. gondii MGIS4-8
27	21.8	34.1	2417	20	X91431	T. gondii MGIS4-8
28	21.6	33.8	2059	19	V52041	Helicobacter poly
29	21.4	33.4	1551	21	Z98176	Human signal pepti
30	21.4	33.4	2097	19	Z96306	S. pneumoniae derl
31	21.4	33.4	2224	20	X10617	cdNA encoding a mu
32	21.4	33.4	2948	20	X18940	Human basic helix-
33	21.4	33.4	3071	19	V65214	DNA encoding a S.
34	21.4	33.4	4060	20	X20655	Polynucleotide seq
35	21.4	33.4	9927	20	X12997	Enterococcus faeca
36	21.4	33.4	20199	19	V52139	Streptococcus pneu
37	21.4	33.4	49999	20	Z23899	Human LOBO homolog
38	21.2	33.1	873	19	V47586	Leishmania antigen
39	21.2	33.1	2277	19	V13836	Homo sapiens mamma
40	21.2	33.1	2277	19	V05372	Human telomerase p
41	21.2	33.1	7198	20	X60278	DNA encoding a 2,4
42	21.2	33.1	99960	21	Z50905	Human TBC-1 parlia
43	21	32.8	340	20	X41055	Human secreted pro
44	21	32.8	598	21	Z80454	Human colon cancer
45	21	32.8	607	21	Z80335	Human colon cancer

ALIGNMENTS

RESULT 1

T42852

ID T42852 standard; DNA: 173 BP.

AC T42852:

XX

XX

DT 16-JUN-1997 (first entry)

DE Promoter #2 for phospholipase D gene.

DE

XX

XX

KW Phospholipase D; rice; promoter; ss.

KW

XX

OS Oryza sativa.

OS

XX

XX

PN W09630510-A1.

PN

XX

PD 03-OCT-1996.

PD

XX

XX

PE 28-MAR-1996; 96MO-JP00812.

PE

XX

XX

PR 29-MAR-1995; 95JP-0096126.

PR

XX

XX

PA (NISB) JAPAN TOBACCO INC.

PA

XX

XX

PI Morioka S, Ueki J;

PI

XX

DR WPI; 1996-455357/45.

DR

XX

XX

PT Promoter DNA sequence derived from rice - used to increase

PT

XX

XX

PS expression of foreign genes in transformed hosts

PS

XX

XX

PS Claim 3; Page 23; 29pp; Japanese.

PS

CC T42851 and T42852 represent promoters isolated from rice. These sequences

CC

CC

are specifically promoters for the phospholipase D gene (PLD). This

CC sequence represents a shortened version of the promoter sequence shown in
CC T42851. These sequences are efficient promoters for greatly increasing
CC the expression of foreign genes in transformant rice and other plants.
XX
SQ Sequence 173 BP; 21 A; 48 C; 47 G; 57 T; 0 other;

Query Match 100.0%; Score 64; DB 17; Length 173;
Best Local Similarity 100.0%; Pred. No. 9.1e-15;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgctgcttaagcgcactagagctctctgctgctctctccg 60
|||||
DB 2 taagccagtgctgcttaagcgcactagagctctctgctgctctctccg 61

OY 61 ctca 64
||||
DB 62 ctca 65

RESULT 2
V05554
ID V05554 standard; DNA; 173 BP.

AC V05554;
DT 18-JUN-1998 (first entry)
DE Ubiquitin gene Intron.

XX
XX Intron; Oryza sativa; ubiquitin gene; ds.
KM
XX

OS Oryza sativa.
XX
XX WO9747755-A1.
PN

PD 18-DEC-1997.
XX
XX 12-JUN-1997; 97WO-JP02030.
PF

XX 12-JUN-1996; 96JP-0172922.
PR
XX (NISR) JAPAN TOBACCO INC.
PA

XX Kuraya Y, Morioka S, Ohta S, Ueki J;
PI
XX WPI; 1998-086573/08.
DR

PT Method for expressing foreign gene - comprises inserting gene
PT downstream from promoter comprising at least two introns, useful in,
PT e.g. genetic engineering
PS

PS Claim 4; Page 10; 20pp: Japanese.
XX

CC This sequence represents an Intron from the Oryza sativa ubiquitin gene,
CC and can be used in the method of the invention. The method is for
CC expressing a foreign gene by inserting the gene downstream from a
CC promoter, comprising at least 2 intron sequences. The method is used for
CC expression of foreign genes, useful in, e.g. genetic engineering. The
CC method achieves higher levels of expression than conventional methods.
XX

SQ Sequence 173 BP; 21 A; 48 C; 47 G; 57 T; 0 other;

Query Match 100.0%; Score 64; DB 19; Length 173;
Best Local Similarity 100.0%; Pred. No. 9.1e-15;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgctgcttaagcgcactagagctctctgctgctctctccg 60
|||||
DB 2 taagccagtgctgcttaagcgcactagagctctctgctgctctctccg 61

OY 61 ctca 64

DB 62 ctca 65

RESULT 3

T42851
ID T42851 standard; DNA; 183 BP.

XX
AC T42851;
XX

DT 16-JUN-1997 (first entry)
DE

XX Promoter #1 for phospholipase D gene.
XX

KM Phospholipase D; rice; promoter; ss.
XX

OS Oryza sativa.
XX

PN WO9630510-A1.
PD

PD 03-OCT-1996.
XX

PF 28-MAR-1996; 96WO-JP00812.
XX

PR 29-MAR-1995; 95JP-0096126.
XX

PA (NISR) JAPAN TOBACCO INC.
XX

PI Morioka S, Ueki J;
XX

DR WPI; 1996-455357/45.
XX

PT Promoter DNA sequence derived from rice - used to increase
PT expression of foreign genes in transformed hosts
XX

PS Claim 1; Page 14; 29pp: Japanese.
XX

CC T42851 and T42852 represent promoters isolated from rice. These sequences
CC are specifically promoters for the phospholipase D gene (PLD). T42852
CC represents a shortened version of this sequence. These sequences are
CC efficient promoters for greatly increasing the expression of foreign
CC genes in transformant rice and other plants.
XX

SQ Sequence 183 BP; 23 A; 51 C; 51 G; 58 T; 0 other;

Query Match 100.0%; Score 64; DB 17; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taagccagtgctgcttaagcgcactagagctctctgctgctctctccg 60
|||||
DB 7 taagccagtgctgcttaagcgcactagagctctctgctgctctctccg 66

OY 61 ctca 64
||||
DB 67 ctca 70

RESULT 4
Q86785
ID Q86785 standard; DNA; 2799 BP.

XX
AC Q86785;
XX

DT 04-MAR-1996 (first entry)
DE

XX DNA encoding Phospholipase D.
XX

KM phospholipase D; measurement; reagent; phospholipid level; ss.
XX

OS Zea mays.
XX

```

FH Key Location/Qualifiers
FT exon 1876..1983
FT /*tag= a
FT 1984..2523
FT intron /*tag= b
FT 2524..2799
FT exon /*tag= c
FT /note= "partial exon 2"

PN W09509234-A1.
XX
XX 06-APR-1995.
XX
XX 30-SEP-1994; 94WO-JP01627.
XX
XX 30-SEP-1993; 93JP-0267884.
XX
XX (NISB ) JAPAN TOBACCO INC.
XX
XX Morioka S, Ueki J;
XX WPI: 1995-147433/19.
XX P-PSDB: R72799.
XX
XX Cloned DNA coding plant derived phospholipase D - controls
XX expression of plant derived PLD gene
XX
XX Claim 14; Page 33-35; 41pp: Japanese.
XX
XX The DNA contains sequences necessary for the expression of a plant
XX derived phospholipase D (PLD). The PLD is useful for measuring
XX phospholipid levels and for producing derivs. by e.g. base exchange
XX reactions.
XX
SQ Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;

Query Match 100.0%; Score 64; DB 16; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.7e-14;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taagccagtgctgcttaagcctaagcgacactgcttcgtcgtctctccg 60
    |||||||
Db 1667 taagccagtgctgcttaagcctaagcgacactgcttcgtcgtctctccg 1726

OY 61 ctca 64
    ||||
Db 1727 ctca 1730

RESULT 5
T42854
ID T42854 standard; cDNA to mRNA; 2799 BP.
XX
XX T42854;
XX
XX 16-JUN-1997 (first entry)
XX
XX Phospholipase D gene sequence fragment.
XX
XX Phospholipase D; rice; promoter; ss.
XX
XX Oryza sativa.
XX
XX
FH Key Location/Qualifiers
FT exon 1876..1983
FT /*tag= a
FT /number= 1
FT /note= "encodes residues 1 to 36 of W06134"
FT intron /*tag= b
FT /number= 1
FT 2524..2799
FT exon
```

```

FT /*tag= c
FT /number= 2
FT /note= "encodes residues 37 to 128 of W06134"

PN W09630510-A1.
XX
XX 03-OCT-1996.
XX
XX 28-MAR-1996; 96WO-JP00812.
XX
XX 29-MAR-1995; 95JP-0096126.
XX
XX (NISB ) JAPAN TOBACCO INC.
XX
XX Morioka S, Ueki J;
XX WPI: 1996-455357/45.
XX P-PSDB: W06134.
XX
XX Promoter DNA sequence derived from rice - used to increase
XX expression of foreign genes in transformed hosts
XX
XX Disclosure; Page 20-22; 29pp: Japanese.
XX
XX This sequence represents a fragment of the coding sequence of the rice
XX phospholipase D gene (PLD). The promoter for the PLD gene was isolated
XX using the primers shown in T42857 and T42588. The promoters (see T42851
XX and T42852) are efficient promoters for greatly increasing the expression
XX of foreign genes in transformant rice and other plants.
XX
SQ Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;

Query Match 100.0%; Score 64; DB 17; Length 2799;
Best Local Similarity 100.0%; Pred. No. 1.7e-14;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taagccagtgctgcttaagcctaagcgacactgcttcgtcgtctctccg 60
    |||||||
Db 1667 taagccagtgctgcttaagcctaagcgacactgcttcgtcgtctctccg 1726

OY 61 ctca 64
    ||||
Db 1727 ctca 1730

RESULT 6
N70891
ID N70891 standard; DNA; 1331 BP.
XX
XX N70891;
XX
XX 26-FEB-1991 (first entry)
XX
XX DNA sequence encoding tumor necrosis factor.
XX
XX Tumor necrosis factor; ss; antitumor.
XX
XX
FH Key Location/Qualifiers
FT mat_peptide 424..885
FT /*tag= a
FT /label= tumor necrosis factor protein

PN JP62135493-A.
XX
XX 18-JUN-1987.
XX
XX 06-DEC-1985; 85JP-0275392.
XX
XX 06-DEC-1985; 85JP-0275392.
XX
XX (TANA ) TANABE SEIYAKU KK.
XX
```

DR		WPI: 1987-209318/30.
DR	P-PSDB; P70538.	
XX		
PT	Anticancer factor - expressed by specified amino acid sequence.	
XX		
PS	Disclosure: Page 11-12; 17pp; Japanese.	
CC	The DNA sequence encodes tumor necrosis factor which is a lymphokine	
CC	with antitumor activity.	
CC	See also P70557.	
XX		
SQ	Sequence 1331 BP; 318 A; 403 C; 326 G; 284 T; 0 other:	
OY	Query Match	38.8%; Score 24.8; DB 8; Length 1331;
	Best Local Similarity	67.3%; Pred. No. 3.3;
	Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;	
DJ	11 gtcctagcctaagcgacactagatcttctgctgcgtctctccgcgc 62	
	1227 gggcttcgtttctctcgtctctgctctctgctctcttgccttacacgc 1278	
RESULT 7		
ID	T99573 standard; DNA: 2110 BP.	
AC	T99573;	
DT	08-JUN-1998 (first entry)	
DE	Mouse Friend virus susceptibility 1 gene (Fvln allele).	
KX	Friend virus susceptibility gene; FvI; FvIn; mouse; retrovirus;	
KM	Infection; antiviral; virulence; ds.	
OS	Mus musculus strain AKR/J.	
FH		
FT	Key Location/Qualifiers	
CDS	35..1357	
FT	/*tag= a	
PM	WO9743410-A1.	
PD	20-NOV-1997.	
PF	14-MAY-1997; 97MO-GB01311.	
PR	14-MAY-1996; 96GB-0009995.	
PA	(MEDIC-) MEDICAL RES COUNCIL.	
PI	Le Tissier PR; Stoye JP;	
DR	WPI: 1998-008879/01.	
DR	P-PSDB; W26782.	
PT	Murine Friend virus susceptibility 1 gene polypeptide - which	
PT	inhibits retroviral integration into newly infected cells, useful in	
XX	antiviral agent development	
XX		
PS	Claim 6; Page 41-45; 55pp; English.	
CC	This polynucleotide comprises the mouse friend virus susceptibility	
CC	1 gene allele Fvln. It was isolated from an EcoRI library prepared	
CC	in the lambda vector DMSH II by screening with a 190 bp fragment,	
CC	synchronised by PCR (see T99578-79), which maps to a cosmid clone	
CC	prepared from C57BL6/J yeast artificial chromosome D11. Another	
CC	FvI allele, FvIb (see T99572), was isolated from C57BL6/J mice.	
CC	The FvI gene map position on chromosome 4 is 76.5 CM. FvI	
CC	transcripts can not be detected by Northern blotting, suggesting	
CC	very low levels of in vivo expression. Claimed host cells can be	
CC	used for the recombinant production of FvI polypeptide. The	

CC	polypeptide inhibits retroviral integration into newly infected
CC	cells. Fragments of the polypeptide can be used as antiviral agents.
CC	Also contemplated is use of the nucleic acid sequence in gene
CC	therapy, and to screen for modulators of the polypeptide or modified
CC	forms of the polypeptide that inhibit retroviral replication.
XX	
SO	Sequence 2110 BP; 558 A; 446 C; 527 G; 579 T; 0 other;
<hr/>	
Query Match	38.8%; Score 24.8; DB 19; Length 2110;
Best Local Similarity	67.3%; Pred. No. 3.7; Mismatches 17; Indels 0; Gaps
Matches 35; Conservative 0;	
OY	11 ggtcctagacctaaggcactagacctcttgccttgcttcctccgct 62 1951 gggttcgtttctctctcctccttgcttccttgcttccttaacgct 2002
Db	
RESULT 8	
X30150	
ID	X30150 standard; DNA; 1037 BP.
AC	X30150;
DT	18-JUN-1999 (first entry)
XX	
DE	Human secreted protein gene 6.
XX	
KW	Human; secreted protein; cancer; tumour; developmental abnormality;
KW	fetal deficiency; blood disorder; immune system disorder; inflammation;
KW	autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW	scleroderma; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW	atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW	digestive disorder; endocrine disorder; infection; AIDS; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9910363-A1.
XX	
PD	04-MAR-1999.
XX	
PF	27-AUG-1998; 98MO-US17709.
XX	
PR	29-AUG-1997; 97US-0056271.
PR	29-AUG-1997; 97US-0056073.
PR	29-AUG-1997; 97US-0056247.
PR	29-AUG-1997; 97US-0056270.
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
XX	
DR	MP1: 1999-190585/16.
DR	P-PSDB; Y04298.
XX	
PT	New isolated human genes and the secreted polypeptides they encode
PT	- useful for diagnosis and treatment of e.g. cancers, neurological
PT	disorders, immune diseases, inflammation or blood disorders
XX	
XX	Claim 1; Page 135; 170pp; English.
XX	
CC	X30145 to X30173 represent 29 isolated human secreted protein genes.
CC	X04293 to X04321 represent the secreted proteins encoded by the 29 human
CC	genes. The genes and their corresponding secreted polypeptides are
CC	useful for preventing, treating or ameliorating medical conditions,
CC	e.g. by protein or gene therapy. Also pathological conditions can be
CC	diagnosed by determining the amount of the new polypeptides in a sample
CC	or by determining the presence of mutations in the new genes. Specific
CC	uses are described for each of the 29 genes, based on which tissues they
CC	are most highly expressed in, and include developing products for the
CC	diagnosis or treatment of cancer, tumours, developmental abnormalities
CC	and foetal deficiencies, blood disorders, diseases of the immune system,
CC	autoimmune diseases, inflammation allergies, Alzheimer's and cognitive
CC	

CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
 CC disorders, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The sequences given in X30174 to X30182 and Y04322 to Y04334 are used
 CC in the exemplification of the present invention.

XX Sequence 1037 BP; 308 A; 198 C; 224 G; 306 T; 1 other;

Query Match 36.9%; Score 23.6; DB 20; Length 1037;

Best Local Similarity 69.6%; Pred. No. 8.6; Mismatches 14; Gaps 0;

Y 18 ggcctagcgcactagctcttgcctgccttccttcctcgcgc 63
 |||| | |||| | || |||| | |||| | |||| | ||||
 Db 14 ggcctgtggcactggaactagctgctgctacatcaagtc 59

RESULT 9

ID Q05873 standard; DNA; 2306 BP.

AC Q05873;

DT 07-JAN-1991 (first entry)

DE Sequence encoding murine interleukin-2 receptor beta-chain.

KM IL-2R beta: pIL-2R beta 9; ss;

OS Mus musculus.

FH Key Location/Qualifiers

FT CDS 375..1994

FT /*tag= a

XX EP386304-A.

PD 12-SEP-1990.

PF 29-MAY-1989; 89EP-0109656.

PR 29-MAY-1989; 89EP-0109656.

PR 07-MAR-1989; 89EP-0104023.

PA (OSAU) OSAKA UNIVERSITY.

PI Taniguchi T;

PI WPI: 1990-276456/37.

DR P-PSDB: R06647.

XX Recombinant interleukin-2 receptor beta chain - used for studying

PT IL-2 system and producing antibodies for diagnosis and therapy.

XX Claim 6; Fig 8; 37pp; English.

XX IL-2 receptor beta chains are useful in studying the biochemistry of

CC interleukin, and in diagnosis and therapy by immune suppression and

CC activation.

XX Sequence 2306 BP; 482 A; 688 C; 585 G; 551 T; 0 other;

Query Match 36.9%; Score 23.6; DB 11; Length 2306;
 Best Local Similarity 76.3%; Pred. No. 10;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Y 25 cgcactagactcttgcctgccttccttccttcgcgc 62
 | | | | | | | | | | | | | | | | | | | | | |
 Db 9 ctctcttgccttgccttccttccttccttcacagct 46

RESULT 10
 ID Q06332 standard; cDNA; 2306 BP.

AC Q06332;

DT 07-FEB-1991 (first entry)

DE Murine IL-2R beta chain.

KM Interleukin; receptor; ss.

OS Mus musculus.

FH Key Location/Qualifiers

FT CDS 375..1994

FT misc_difference 1182..1184

FT /*tag= a

FT /*tag= b

FT /label=ggc, tgc

XX AU9050726-A.

PD 13-SEP-1990.

PF 06-MAR-1990; 90AU-0050726.

PR 20-JUL-1989; 89EP-0113310.

PR 07-MAR-1989; 89EP-0104023.

PR 29-MAY-1989; 89EP-0109656.

XX (BOEH) BOEHRINGER INGELHEI.

XX Taniguchi T, Hatakeyama M, Minamoto S, Kono T, Doi T;

PI Miyasaka M, Tsudo, Karasuyama H;

DR WPI: 1990-327673/44.

DR P-PSDB: R07507.

XX Recombinant interleukin-2 receptor beta chain - useful for

PT diagnosis and therapy by immune suppression or activation.

XX Claim 4; Fig 8; 65pp; English.

XX The sequence was obtd. from clones isolated from a cDNA library

CC prepd. from RNA extracted from Concanavalin A-stimulated mouse

CC spleen cells. The clones were selected using a probe derived from

CC the human sequence. The sequence encodes a soluble portion of the

CC IL-2R beta chain.

CC See also Q06331.

XX Sequence 2306 BP; 487 A; 685 C; 580 G; 550 T; 4 other;

Query Match 36.9%; Score 23.6; DB 11; Length 2306;
 Best Local Similarity 76.3%; Pred. No. 10;
 Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Y 25 cgcactagactcttgcctgccttccttccttcgcgc 62

| | | | | | | | | | | | | | | | | | | | | |

Db 9 ctctcttgccttgccttccttccttccttcacagct 46

RESULT 11

X13021/c

ID X13021 standard; DNA; 7947 BP.

AC X13021;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome config SEQ ID NO:84.

XX	Enterococcus faecalis; confliq; detection: Enterococcal Infection;	
KW	vaccine; attenuation; computer readable medium; ds.	
XX		
OS	Enterococcus faecalis.	
PN		
XX	MO9850555-A2.	
PD		
XX	12-NOV-1998.	
XX		
PF	04-MAY-1998; 98WO-US08985.	
XX		
PR	14-NOV-1997; 97US-0066009.	
PR	06-MAY-1997; 97US-0044031.	
PR	16-MAY-1997; 97US-0046655.	
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
P1	Barash SC, Dillon PJ, Kunsch CA;	
XX		
DR	WPI: 1999-045171/04.	
XX		
PT	New Isolated Enterococcus faecalis polynucleotides and polypeptides	
PT	- used to develop products for the detection of Enterococcus and for	
PT	use in vaccines for prevention or attenuation of Enterococcus	
PT	infection.	
XX		
PS	Claim 1: Page 580-584; 2084pp; English.	
CC		
CC	A computer readable medium has been developed which has recorded on it	
CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome.	
CC	X12938 to X13919 represent these nucleotide sequences which are primarily	
CC	nucleotide sequences, also known as confliqs. The computer-based system	
CC	can identify fragments of the Enterococcus faecalis genome with	
CC	commercial importance. The products can be used to detect the presence	
CC	of Enterococcus faecalis in samples. They can also be used for	
CC	diagnosing Enterococcal infection in an animal and monitoring	
CC	progression of disease, and for identifying agents which can be used to	
CC	modulate the growth or pathogenicity of Enterococcus faecalis, or	
CC	another related organism, in vivo or in vitro. In particular the	
CC	polypeptides encoded by the Enterococcus faecalis nucleotide sequences	
CC	can be used in vaccines to prevent or attenuate an Enterococcal	
CC	infection.	
XX		
XX	Sequence 7947 BP; 2720 A; 1379 C; 1741 G; 2100 T; 7 other;	
QY		
Db	Query Match 36.9%; Score 23.6; DB 20; Length 7947;	
	Best Local Similarity 69.6%; Pred. No. 14;	
	Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;	
	8 agtgccttagcgtacgaacacctaagcttcctgcctgcctc 53	
	4084 AGTAGCTTCGCTAACGCTGTCTTGGCTTCCTGATCCGCTTCTC 4039	
RESULT 12		
T87466/c		
ID T87466	standard; DNA; 508 BP.	
XX T87466;		
XX		
DT 29-JAN-1998	(first entry)	
XX		
DE	Hamster Ubiquitin/s27a cDNA sequence.	
XX		
KW	CHO; hamster; Ub/S27a; promoter; ubiquitin; primer; ss.	
XX		
OS	Hamster.	
XX		
FIH	Key	Location/Qualifiers
FT	CDS	17..487
FT		/*tag= a

FT	misc_RNA	1..244
FT	/*tag= b	
FT	/note= "ubiquitin"	
FT	intron	64..65
FT	/*tag= c	
FT	/note= "Site of intron"	204..205
FT	/*tag= d	
FT	/note= "Site of intron"	337..338
FT	/*tag= e	
FT	/note= "Site of intron"	497..502
FT	/*tag= f	
FT	polyA_signal	501..506
FT	/*tag= g	
XX	DE19539493-A1.	
XX	30-APR-1997.	
XX	24-OCT-1995;	95DE-1039493.
XX	24-OCT-1995;	95DE-1039493.
XX	(THOM) THOMAE GMBH KARL.	
XX	Bergemann K, Enekel B, Gannon F, Noe W;	
XX	WPI: 1997-246232/23.	
DR	P-PSDB: W28542.	
XX	Nucleic acid containing promoter and regulatory regions of the	
PT	hamster ubiquitin S27a gene - useful for production of heterologous	
PT	proteins, especially in CHO cells	
XX	Claim 30; Fig 1; 22pp; German.	
PS	The cDNA sequences of Ub/S27a isolated from CHO cells (T87466)	
CC	and from human cells (T87467) were compared. They show 92.2%	
CC	homology. The corresponding amino acid sequences show 100%	
CC	homology.	
SO	Sequence 508 BP; 159 A; 98 C; 130 G; 121 T; 0 other;	
Query Match	36.6%; Score 23.4; DB 18; Length 508;	
Best Local Similarity	67.3%; Pred. No. 8.6; Mismatches 16; Indels 0; Gaps 0;	
Matches 33; Conservative 0;		
QY	11 ggtcctaagctaagcgcacactagagcttcgtcctgcgtctctctcc 59	
Db	331 gTACTTGACGACAGCCACTTACCTTCTTCTTATGCTTATTTCTTC 283	
RESULT 13		
Z20473		
ID	Z20473 standard; DNA: 219 BP.	
XX	Z20473;	
XX	19-NOV-1999 (first entry)	
DE	Barley microsatellite marker clone Bmag5.	
XX	Microsatellite marker; barley; chromosome 7 marker; cereal;	
KW	fermentability; group 5 chromosome; ethyl carbanate production; Bmac213;	
KW	wort fermentation; Triticaceae; Bmac96; epi-heterodendrin production;	
XX	diagnosis; ss.	
CS	Hordeum vulgare.	
PN	W09946404-A1	

XX 16-SEP-1999.
 PD
 XX 01-MAR-1999; 99WO-GB00602.
 XX
 XX 10-MAR-1998; 98GB-0005087.
 PR
 XX (SCCR-) SCOTTISH CROP RES INST.
 PA
 XX Thomas WTB, Swanson JS, Powell W, Maugh R, Ramsey LD;
 PI
 XX WPI: 1999-551424/46.
 DR
 XX
 XX Screening cereals for fermentability, especially useful in barley -
 PT
 XX
 XX Disclosure: Fig 7; 49pp; English.
 PS
 XX This sequence represents a barley microsatellite marker. Primers that
 CC amplify this sequence can be used in the method of the invention.
 CC The method is for screening cereal for fermentability, comprising
 CC analysing cereal genomic DNA to determine which allele(s) of a gene/gene
 CC complex affecting fermentability at a locus close to the centromere on
 CC homologous Triticeae group 5 chromosome (barley chromosome 7) is/are
 CC present. The invention also relates to a method for screening cereal for
 CC ethyl carbamate production on wort fermentation and distillation.
 CC comprising analysing barley genomic DNA to determine which allele(s) of
 CC the locus, designated eph on the short arm of homologous Triticeae
 CC group 1 chromosome (barley chromosome 5) is/are present. The methods and
 CC primers are useful for identifying microsatellites Bmac96 and Bmac213,
 CC which are useful for determining fermentability and/or epi-heterodendrin
 CC production in cereals, especially barley. Current methods for determining
 CC fermentability are difficult to apply within barley breeding programs.
 CC Prior art methods using molecular markers have difficulty in detecting
 CC levels of allelic variation.
 CC
 SQ Sequence 219 BP; 42 A; 48 C; 45 G; 84 T; 0 other:

Query Match 35.6%; Score 22.8; DB 20; Length 219;
 Best Local Similarity 66.0%; Pred. No. 12;
 Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 9 gtgtgcttagcgaagcactagctcttcgtcgtcttcttc 58
 ||||| ||| | | | | |||| | ||| | |||| |||||
 DB 51 gtgtgcatagacctacggcgtccatagctagctagctattcttc 100

RESULT 14
 V45981
 ID V45981 standard; DNA: 2353 BP.
 XX
 AC V45981;
 XX
 DT 09-OCT-1998 (first entry)
 XX
 XX A. thaliana sterol glycosyltransferase cDNA.
 DE
 XX Sterol glycosyl transferase; oat; transgenic plant; tolerance;
 KM resistance; environmental factor; drought; salt; cold; frost;
 KM fungal attack; temperature; secondary metabolite; inhibitor; antitumour;
 KM vascular permeability; anti-inflammatory; haemostatic agent; ss.
 KW
 XX Arabidopsis thaliana.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 113..2026
 FT /*tag= a
 FT /product= SGT
 FT /note= "sterol glycosyltransferase"
 XX
 XX MO9817789-A1.
 XX
 PD 30-APR-1998.

XX 10-OCT-1997; 97WO-DE02335.
 PF
 XX 21-OCT-1996; 96DE-1043309.
 PR
 XX (BALT/) BALTRUSCH A.
 PA (BALT/) BALTRUSCH R M.
 PA (GVSE-) GVS GES ERWERB & PA VERWERTUNG LANDWIRTSCHA.
 XX
 XX Baltrusch M, Heinz E, Warnecke D, Wolter FP;
 PI
 XX WPI: 1998-271789/24.
 DR
 XX P-PSDB: W64388, W64389
 DR
 XX
 XX DNA encoding sterol glycosyltransferase and related enzymes -
 PT transgenic plants, with increased resistance to stress, and
 PT organisms, used to produce steryl glycoside(s), potentially useful
 PT therapeutically, in high yield
 PS
 XX Example 9; Fig 17; 64pp; German.
 CC
 CC This sequence encodes a sterol glycosyltransferase (SGT) isolated from
 CC A. thaliana which is used in a method to produce transgenic plants which
 CC have a better tolerance of/resistance against adverse environmental
 CC factors such as drought, high salt content of the soil, cold/frost and
 CC fungal attack using recombinant SGT. Microorganisms containing a
 CC recombinant SGT protein have better resistance to high salt/ethanol
 CC content of growth media, cold/frost and high temperature. Organisms
 CC containing such recombinant proteins are used to produce secondary
 CC metabolites which are potentially useful for inhibiting vascular
 CC permeability, or as antitumour, anti-inflammatory and haemostatic agents.
 CC
 SQ Sequence 2353 BP; 700 A; 468 C; 518 G; 667 T; 0 other:

Query Match 35.6%; Score 22.8; DB 19; Length 2353;
 Best Local Similarity 71.4%; Pred. No. 20;
 Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 18 ggctagcgacagcagctcttcgtcgtcttcttc 59
 ||||| ||| | | | ||||| |||| | ||||| |||||
 DB 130 ggctgagctgcgcagagttcttcctcgtcttcttcttc 171

RESULT 15
 V03752
 ID V03752 standard; CDNA: 2464 BP.
 XX
 AC V03752;
 XX
 DT 15-APR-1998 (first entry)
 XX
 XX Porcine TNFalpha-convertase coding sequence.
 DE
 XX Tumour necrosis factor alpha convertase; TNFalpha-convertase; human;
 KM inhibitor; therapy; systemic inflammatory response syndrome; malignancy;
 KM reperfusion injury; cardiovascular disease; infectious disease; cachexia;
 KM obstetrical disorder; gynaecological disorder; inflammatory disease; pig;
 KM autoimmunity; allergic disease; atopic disease; septic shock; psoriasis;
 KM graft osteoporosis; restenosis; rheumatoid arthritis; osteoarthritis;
 KM macular degeneration; multiple sclerosis; AIDS; porcine; ss.
 KW
 XX Sus scrofa.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1311
 FT /*tag= a
 FT
 XX
 XX MO9735538-A2.
 XX
 PD 02-OCT-1997.
 XX
 XX 25-MAR-1997; 97WO-EP01497.
 PF

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 20:11:54 ; Search time 1275.75 Seconds

(without alignments)
256,740 Million cell updates/sec

Title: US-09-600-602-1

Perfect score: 64
Sequence: 1 taagcccaagtgtgcttaagc.....gcttgctctctcgcgtca 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_ph: *
6: gb_pl1: *
7: gb_pl2: *
8: gb_pr1: *
9: gb_pr2: *
10: gb_pr3: *
11: gb_ro: *
12: gb_sy: *
13: gb_un: *
14: em_fun: *
15: em_hum1: *
16: em_hum2: *
17: em_in: *
18: em_om: *
19: em_or: *
20: em_ov: *
21: em_pat: *
22: em_ph: *
23: em_pl: *
24: em_ro: *
25: em_sts: *
26: em_sy: *
27: em_un: *
28: em_vl: *
29: gb_htg1: *
30: gb_htg2: *
31: gb_in1: *
32: gb_in2: *
33: em_ba1: *
34: em_ba2: *
35: em_hum3: *
36: em_hum4: *
37: gb_pr4: *
38: gb_htg3: *
39: gb_htg4: *
40: gb_htg5: *
41: gb_htg6: *
42: gb_htg7: *
43: em_htg1: *

44: em_htg2: *
45: em_htg3: *
46: em_hum5: *
47: gb_pl3: *
48: gb_pr5: *
49: gb_htg8: *
50: gb_htg9: *
51: gb_htg10: *
52: gb_htg11: *
53: gb_htg12: *
54: gb_htg13: *
55: gb_htg14: *
56: gb_in3: *
57: gb_htg15: *
58: gb_htg16: *
59: gb_htg17: *
60: em_htg4: *
61: em_htg5: *
62: em_htg6: *
63: em_htg7: *
64: em_hum6: *
65: gb_htg18: *
66: gb_htg19: *
67: gb_htg20: *
68: gb_htg21: *
69: gb_htg22: *
70: gb_htg23: *
71: gb_vil: *
72: gb_vil2: *
73: gb_ba3: *
74: em_htg8: *
75: em_htg9: *
76: em_htg10: *
77: gb_pr6: *
78: gb_pr7: *
79: gb_sts1: *
80: gb_sts2: *
81: gb_pat1: *
82: gb_pat2: *
83: em_htg0: *
84: gb_htg24: *
85: gb_pr8: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	64	100.0	173	81	AR037064	AR037064 Sequence
2	64	100.0	183	81	AR037061	AR037061 Sequence
3	64	100.0	2799	81	AR005013	AR005013 Sequence
4	64	100.0	2799	81	AR037063	AR037063 Sequence
5	64	100.0	5871	6	AB001920	AB001920 Oryza sat
6	28	43.8	161234	48	CNS01DUE	AL133241 Human chr
7	26.4	41.2	87974	49	AC021283	AC021283 Homo sapi
8	26.4	41.2	162866	50	AC023277	AC023277 Homo sapi
9	26.4	41.2	173556	59	AC074388	AC074388 Homo sapi
10	26.4	41.2	207550	40	AC015977	AC015977 Homo sapi
11	26.4	41.2	215708	57	AC068757	AC068757 Homo sapi
12	26.2	40.9	6036	11	RATPDC	LO1115 Rattus norv
13	26	40.6	110000	84	HSXK1_1	Continuation (2 of
14	26	40.6	110000	84	HSXK1_2	Continuation (3 of
15	26	40.6	300050	78	HSXK1A	AL442166 Homo sapi
16	26	40.6	340000	77	HS21C084	AL163284 Homo sapi
17	25.8	40.3	958	79	CNS01M44	AL150993 Anopheles
18	25.6	40.0	76141	51	AC024393	AC024393 Homo sapi
19	25.6	40.0	106784	66	AL139250	AL139250 Homo sapi
20	25.6	40.0	168982	69	AL390994	AL390994 Homo sapi
21	25.4	39.7	21721	32	CEP16H9	250005 Caenorhabdi

```
22 25.4 39.7 172551 9 AC008518
C 23 25.4 39.7 176616 51 AC026427
C 24 25.4 39.7 186340 41 AC016684
25 25.2 39.4 3893 6 AF036763
26 25.2 39.4 5626 6 AF043332
C 27 25.2 39.4 111810 50 AC022139
C 28 25.2 39.4 147880 49 AC021945
29 25.2 39.4 152290 40 AC016107
C 30 25.2 39.4 169069 30 AC008871
C 31 25.2 39.4 179686 51 AC023987
C 32 25.2 39.4 186603 54 AC036155
C 33 25.2 39.4 192638 41 AC017063
34 25.2 39.4 206074 10 AC009484
35 25.2 39.1 41193 31 AC013257
C 36 25 39.1 116215 8 AC004130
C 37 25 39.1 172083 10 AC025614
C 38 25 39.1 219194 49 AC022061
C 39 24.8 38.8 370 11 MUSCMOSR2
C 40 24.8 38.8 539 11 MUSCMOS
41 24.8 38.8 1331 81 E01225
42 24.8 38.8 2110 11 MMEFV1A
43 24.8 38.8 2110 81 A67510
44 24.8 38.8 88028 8 AC004938
C 45 24.8 38.8 97832 78 HSDJ46801
```

ALIGNMENTS

```
RESULT 1
LOCUS AR037064 173 bp DNA
DEFINITION Sequence 7 from patent US 5801016.
ACCESSION AR037064
VERSION AR037064.1 GI:5954920
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 173)
AUTHORS Morloka,S. and Ueki,J.
TITLE DNA fragment, recombinant vector containing the same and method for
JOURNAL expressing foreign genes using the same
FEATURES
source location/Qualifiers
1..173
BASE COUNT 21 a 48 c 47 g 57 t
ORIGIN
```

Query Match 100.0%; Score 64; DB 81; Length 173;
Best Local Similarity 100.0%; Pred. No. 8.7e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 taagccagtgctgcttaagcgaactagagctcttgctgcttcttcgcg 60
Db 2 TAAACCCAGTGTCTTAGGCTAAGCGACTAGAGCTTCTGCTGCTTCTCCG 61
```

```
OY 61 ctca 64
Db 62 CTCA 65
```

```
RESULT 2
LOCUS AR037061 183 bp DNA
DEFINITION Sequence 1 from patent US 5801016.
ACCESSION AR037061
VERSION AR037061.1 GI:5954917
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
```

```
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 183)
TITLE DNA fragment, recombinant vector containing the same and method for
JOURNAL expressing foreign genes using the same
FEATURES
source location/Qualifiers
1..183
BASE COUNT 23 a 51 c 51 g 58 t
ORIGIN
```

Query Match 100.0%; Score 64; DB 81; Length 183;
Best Local Similarity 100.0%; Pred. No. 8.7e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 taagccagtgctgcttaagcgaactagagctcttgctgcttcttcgcg 60
Db 7 TAAACCCAGTGTCTTAGGCTAAGCGACTAGAGCTTCTGCTGCTTCTCCG 66
```

```
OY 61 ctca 64
Db 67 CTCA 70
```

```
RESULT 3
LOCUS AR05013 2799 bp DNA
DEFINITION Sequence 5 from patent US 5747327.
ACCESSION AR05013
VERSION AR05013.1 GI:3965892
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2799)
AUTHORS Ueki,J. and Morloka,S.
TITLE Phospholipase D gene originated from plant
JOURNAL Patent: US 5747327-A 5 05-MAY-1998;
FEATURES
source location/Qualifiers
1..2799
BASE COUNT 692 a 709 c 609 g 789 t
ORIGIN
```

Query Match 100.0%; Score 64; DB 81; Length 2799;
Best Local Similarity 100.0%; Pred. No. 9.6e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 taagccagtgctgcttaagcgaactagagctcttgctgcttcttcgcg 60
Db 1667 TAAACCCAGTGTCTTAGGCTAAGCGACTAGAGCTTCTGCTGCTTCTCCG 1726
```

```
OY 61 ctca 64
Db 1727 CTCA 1730
```

```
RESULT 4
LOCUS AR037063 2799 bp DNA
DEFINITION Sequence 4 from patent US 5801016.
ACCESSION AR037063
VERSION AR037063.1 GI:5954919
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2799)
AUTHORS Morloka,S. and Ueki,J.
TITLE DNA fragment, recombinant vector containing the same and method for
```

expressing foreign genes using the same
Patent: US 5801016-A 4 01-Sep-1998;
Location/Qualifiers
1. .2799
/organism="unknown"
BASE COUNT 692 a 709 c 609 g 789 t
ORIGIN

Query Match 100.0%; Score 64; DB 81; Length 2799;
Best Local Similarity 100.0%; Pred. No. 9.6e-13;
Matches .64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taagccagtgcttaagcctaagcagcactagagcttcgtcgtctcttcgcg 60
|||||
Db 1667 TAAGCCAGTGTGCTTAGCGTAAGCGCAGCTTCTGCTGCTTCTCTCG 1726
|||||

Qy 61 ctca 64
|||||
Db 1727 CTCA 1730

RESULT 5
AB001920 5871 bp DNA PLN 14-APR-2000
LOCUS Oryza sativa DNA for phospholipase D, complete cds.
DEFINITION AB001920
ACCESSION AB001920
VERSION 1 GI:1902902
KEYWORDS phospholipase D.
SOURCE Oryza sativa (cultivar:Koshihikari) leaf DNA.
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
1 (bases 152 to 1665; 1839 to 1983; 2524 to 4420)
Ueki,J., Morioka,S., Komari,T. and Kumashiro,T.
Purification and characterization of phospholipase D (PLD) from
rice (Oryza sativa L.) and cloning of cDNA for PLD from rice and
maize (Zea mays L.)
Plant Cell Physiol. 36 (5), 903-914 (1995)
96012933

JOURNAL
MEDLINE 2 (bases 1 to 5871)
REFERENCE Morioka,S., Ueki,J. and Komari,T.
AUTHORS Characterization of two distinctive genomic clones (Accession Nos.
TITLE AB001919 and AB001920) for phospholipase D from rice (PGR97-076)
JOURNAL Plant Physiol. 114, 396 (1997)
REFERENCE 3 (bases 1 to 5871)
AUTHORS Ueki,J.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1995) to the DDBJ/EMBL/GenBank databases. Jun
7001, Japan Tobacco Inc., Plant Breeding and Genetics Research Lab;
700 Higashihara, Iwata, Shizuoka 438-0802, Japan
(E-mail:Jun.ueki@pbgrl.jti.co.jp, Tel:81-538-32-7111,
Fax:81-538-32-8700)
FEATURES
source Location/Qualifiers
1. .5871
/organism="Oryza sativa"
/cultivar="Koshihikari"
/db_xref="taxon:4530"
/tissue_type="leaf"
join(1522. .1665,1839. .1983,2524. .4420,4899. .5702)
/citation=[1]
1522. .1665
/number=1
1666. .1838
/number=1
1839. .1983
/number=2
join(1876. .1983,2524. .4420,4899. .5332)
/EC_number="3.1.4.4"
/citation=[1]
/codon_start=1
/product="phospholipase D"
/protein_id="BAA19467.1"

/db_xref="GI:1902903"
/translation="MAQMLLHGLTLPATIFPAASLSNPNRASGAPKIRKFEVGEIDET
VGWKGATKYVSTIDLEKARYGRTRMTNPEINRWVESPHIYCAHNASVYIFVKID
NIGATNIGRAYLPVQELLNGEELIDICDNRRSEVGSKEIKVQYDYSRDM
ARGYRSTKYRGVPTTFPSQGGCVTLIKQAHVPDNIIPKPLADGKNYEPHRQREI
FDATSNQHLIYITGMSVYETITLVROSNRPKPGDVTTLBELKKASEVRYMLTWM
DIDRTSVGLKRDGLMATHDETEHYFPGSDVNCVLCRNPDGSGIIVQDLSISTMFPH
HOKIVVVDHELIPNGSQORIVSFVGGIDLDGDDYQYHSLEFRTLDSTHDDFHPON
FATASIKKGGPREPMIDHSRLEGPIDMYLYNFEORRQGGKDLLQLRSDTI
PPSPVMPEDPRETMNVLFRSIDGGAAPGFPPEPAKAGIYSGKDOIIDRSTQAY
IHARRAKNTIYINQYFLGSSYAKRPEGIKPEDTGLHLIPKELAKVSKIEAGER
FTYIVVVPMPPEGVESGVQALLDMQRTMEMYTDITLALQAKGLEAPKQYLTFF
CLGNREVKQAGEYQPEQDPEADTDYSRAQEARREMYIVTRKMLVDEYITISANIN
QSRMDGARDESIAMGYQPYHATROPARCOIHGFRLMALVYELHGLMDVDFORPESLE
CVOKVNRILAEKYMDMYSDDLQDLPGLLSYPIGVASDGVYTELPGMEYFPDTRAV
LGAKSDYMPRLITS"
1984. .2523
/number=2
2524. .4420
/number=3
4421. .4898
/number=3
4899. .5702
/number=4
BASE COUNT 1488 a 1363 c 1355 g 1665 t
ORIGIN

Query Match 100.0%; Score 64; DB 6; Length 5871;
Best Local Similarity 100.0%; Pred. No. 9.9e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taagccagtgcttaagcctaagcagcactagagcttcgtcgtctcttcgcg 60
|||||
Db 1667 TAAGCCAGTGTGCTTAGCGTAAGCGCAGCTTCTGCTGCTTCTCTCG 1726
|||||

Qy 61 ctca 64
|||||
Db 1727 CTCA 1730

RESULT 6
CNS01DUE/6 DNA PRI 25-MAY-2000
LOCUS Human chromosome 14 DNA sequence *** IN PROGRESS *** BAC C-2560C21
DEFINITION of library Caltech-D from chromosome 14 of Homo sapiens (human),
complete sequence.
ACCESSION AL133241
VERSION AL133241.3 GI:7630038
KEYWORDS HTG; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 161234)
Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2000) to the EMBL/GenBank/DDBJ databases
AUTHORS On Apr 20, 2000 this sequence version replaced gi:6624578.
COMMENT IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-61A17
Downstream BAC (overlapping the SP6 end) : C-2379B19

Overall quality chart :
Range : bases
0


```

* 25957 26824: contig of 868 bp in length
* 26825 26924: gap of 100 bp
* 26925 27814: contig of 890 bp in length
* 27815 27914: gap of 100 bp
* 27915 28785: contig of 871 bp in length
* 28786 28885: gap of 100 bp
* 28886 29772: contig of 887 bp in length
* 29773 29872: gap of 100 bp
* 29873 30730: contig of 858 bp in length
* 30731 30830: gap of 100 bp
* 30831 31680: contig of 850 bp in length
* 31681 31780: gap of 100 bp
* 31781 32654: contig of 874 bp in length
* 32655 32754: gap of 100 bp
* 32755 33621: contig of 867 bp in length
* 33622 33721: gap of 100 bp
* 33722 34595: contig of 874 bp in length
* 34596 34695: gap of 100 bp
* 34696 35573: contig of 878 bp in length
* 35574 35673: gap of 100 bp
* 35674 36571: contig of 898 bp in length
* 36572 36671: gap of 100 bp
* 36672 37544: contig of 873 bp in length
* 37545 37644: gap of 100 bp
* 37645 38522: contig of 878 bp in length
* 38523 38622: gap of 100 bp
* 38623 39484: contig of 862 bp in length
* 39485 39584: gap of 100 bp
* 39585 40468: contig of 884 bp in length
* 40469 40568: gap of 100 bp
* 40569 41434: contig of 866 bp in length
* 41435 41534: gap of 100 bp
* 41535 42393: contig of 859 bp in length
* 42394 42493: gap of 100 bp
* 42494 43380: contig of 887 bp in length
* 43381 43480: gap of 100 bp
* 43481 44375: contig of 895 bp in length
* 44376 44475: gap of 100 bp
* 44476 45341: contig of 866 bp in length
* 45342 45441: gap of 100 bp
* 45442 46302: contig of 861 bp in length
* 46303 46402: gap of 100 bp
* 46403 47300: contig of 898 bp in length
* 47301 47400: gap of 100 bp
* 47401 48288: contig of 888 bp in length
* 48289 48388: gap of 100 bp
* 48389 49261: contig of 873 bp in length
* 49262 49361: gap of 100 bp
* 49362 50236: contig of 875 bp in length
* 50237 50336: gap of 100 bp
* 50337 51222: contig of 886 bp in length
* 51223 51322: gap of 100 bp
* 51323 52185: contig of 863 bp in length
* 52186 52285: gap of 100 bp
* 52286 53149: contig of 864 bp in length
* 53150 53249: gap of 100 bp
* 53250 54086: contig of 837 bp in length
* 54087 54186: gap of 100 bp
* 54187 55046: contig of 860 bp in length
* 55047 55146: gap of 100 bp
* 55147 55995: contig of 849 bp in length
* 55996 56095: gap of 100 bp
* 56096 56975: contig of 880 bp in length
* 56976 57075: gap of 100 bp
* 57076 57960: contig of 885 bp in length
* 57961 58060: gap of 100 bp
* 58061 58938: contig of 878 bp in length
* 58939 59038: gap of 100 bp
* 59039 59917: contig of 879 bp in length
* 59918 60017: gap of 100 bp
* 60018 60891: contig of 874 bp in length
* 60892 60991: gap of 100 bp
* 60992 61867: contig of 876 bp in length

```

```

* 61868 61967: gap of 100 bp
* 61968 62848: contig of 881 bp in length
* 62849 62948: gap of 100 bp
* 62949 63728: contig of 780 bp in length
* 63729 63828: gap of 100 bp
* 63829 64675: contig of 847 bp in length
* 64676 64775: gap of 100 bp
* 64776 65582: contig of 807 bp in length
* 65583 65682: gap of 100 bp
* 65683 66558: contig of 876 bp in length
* 66559 66658: gap of 100 bp
* 66659 67519: contig of 861 bp in length
* 67520 67619: gap of 100 bp
* 67620 68508: contig of 889 bp in length
* 68509 68608: gap of 100 bp

Query Match      41.2%; Score 26.4; DB 49; Length 87974;
Best Local Similarity 69.2%; Pred. No. 39;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3 agccagtggtgttagtgtaagcgactagagctctctgtctgtcttct 54
DB 66387 ACTGCAGTGGCCCAATCTTGACTCAGTGGACCTGCTCCCGGCTTCT 66438

```

```

RESULT 8
AC023277/c      AC023277 162866 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 4 clone RP11-438E5, WORKING DRAFT SEQUENCE,
DEFINITION 27 unordered pieces.
ACCESSION AC023277
VERSION AC023277.3 GI:7712273
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eularchia; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 162866)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162866)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 6, 2000 this sequence version replaced gi:7024025.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0438E05
----- Summary Statistics -----
Sequencing vector: M13; %
Chemistry: Dye-Primer ET; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151916 bases at least Q40
Consensus quality: 154698 bases at least Q30
Insert size: 160266; sum-of-ctrls
Quality coverage: 3.67 in Q20 bases; agarose-fp
Quality coverage: 3.76 in Q20 bases; sum-of-ctrls

```

* NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

```
* Runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1270: contig of 1270 bp in length
1371: gap of unknown length
1371: contig of 1414 bp in length
2785: gap of unknown length
2885: gap of unknown length
4377: contig of 1493 bp in length
4378: gap of unknown length
4478: contig of 1845 bp in length
6322: contig of 1845 bp in length
6423: gap of unknown length
8110: contig of 1688 bp in length
8111: gap of unknown length
8210: contig of 2174 bp in length
10385: gap of unknown length
10485: contig of 1642 bp in length
12127: gap of unknown length
12226: gap of unknown length
12227: contig of 2641 bp in length
14868: gap of unknown length
14967: gap of unknown length
14968: contig of 3737 bp in length
14968: gap of unknown length
18705: contig of 2077 bp in length
20881: gap of unknown length
20982: contig of 3437 bp in length
24419: gap of unknown length
24519: contig of 3712 bp in length
28231: gap of unknown length
28330: gap of unknown length
32665: contig of 4335 bp in length
32666: gap of unknown length
35962: contig of 3197 bp in length
35963: gap of unknown length
36063: contig of 5277 bp in length
41339: gap of unknown length
41440: contig of 5471 bp in length
46910: gap of unknown length
46911: gap of unknown length
55460: contig of 8450 bp in length
55461: gap of unknown length
55561: contig of 7333 bp in length
62893: gap of unknown length
62993: gap of unknown length
70624: contig of 7631 bp in length
70625: gap of unknown length
78348: contig of 7624 bp in length
78349: gap of unknown length
78449: contig of 7065 bp in length
85513: gap of unknown length
85514: contig of 9214 bp in length
94827: gap of unknown length
94928: contig of 11045 bp in length
105972: gap of unknown length
105973: gap of unknown length
106073: contig of 9948 bp in length
116021: gap of unknown length
116120: contig of 14276 bp in length
116121: gap of unknown length
130396: gap of unknown length
130496: contig of 11873 bp in length
142369: gap of unknown length
142469: contig of 20397 bp in length
142470: location/Qualifiers
142470: location/Qualifiers
1.162866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-438E5"
1.1270
/misc_feature
1371: .2784
/misc_feature
1371: .2784
/misc_feature
2885: .4377
/misc_feature
2885: .4377
/misc_feature
4478: .6322
/misc_feature
4478: .6322
/misc_feature
6423: .8110
/misc_feature
6423: .8110
/note="assembly_name:Contig11"

misc_feature 8211..10384
/note="assembly_name:Contig12"
misc_feature 10485..12126
/note="assembly_name:Contig13"
misc_feature 12227..14867
/note="assembly_name:Contig14"
misc_feature 14968..18704
/note="assembly_name:Contig15"
misc_feature 18805..20881
/note="assembly_name:Contig16"
misc_feature 20982..24418
/note="assembly_name:Contig17"
clone_end:SP6
vector_side:right"
misc_feature 24519..28230
/note="assembly_name:Contig18"
misc_feature 28331..32665
/note="assembly_name:Contig19"
misc_feature 32766..35962
/note="assembly_name:Contig20"
misc_feature 36063..41339
/note="assembly_name:Contig21"
misc_feature 41440..46910
/note="assembly_name:Contig22"
misc_feature 47011..55460
/note="assembly_name:Contig23"
misc_feature 55561..62893
/note="assembly_name:Contig24"
misc_feature 62994..70624
/note="assembly_name:Contig25"
misc_feature 70723..78348
/note="assembly_name:Contig26"
misc_feature 78449..85513
/note="assembly_name:Contig27"
misc_feature 85614..94827
/note="assembly_name:Contig28"
misc_feature 94928..105972
/note="assembly_name:Contig29"
misc_feature 106073..116020
/note="assembly_name:Contig30"
misc_feature 116121..130396
/note="assembly_name:Contig31"
misc_feature 130497..142369
/note="assembly_name:Contig32"
misc_feature 142470..162866
/note="assembly_name:Contig33"

BASE COUNT 49913 a 30224 c 29748 g 50356 t 2625 others
ORIGIN

Query Match 41.2%; Score 26.4; DB 50; Length 162866;
Best Local Similarity 65.0%; Pred. No. 40;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 4 gccacggtgcttaaggctaacgacactgtcttcgtcgttcgttcgttcgcgc 63
||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61109 GCCACGGTTCCTCTCTCTATGTCAGTGGAGCTTGTCCTCTCTGACCTC 61050

RESULT 9
AC074388 173556 bp DNA HTG 13-AUG-2000
LOCUS Homo sapiens chromosome 2 clone RP11-507P19, WORKING DRAFT
DEFINITION SEQUENCE, 21 unordered pieces.
ACCESSION AC074388
VERSION AC074388.2 GI:9690406
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 173556)
AUTHORS Waterston,R.H.
```


TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 173556)
Waterston,R.H.
Direct Submission
Submitted (30-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 4, 2000 this sequence version replaced g1:9587428.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0507P19
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158826 bases at least Q40
Consensus quality: 163318 bases at least Q40
Consensus quality: 165604 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 171556; sum-of-contigs
Quality coverage: 4.60 in Q20 bases; agarose-fp
Quality coverage: 4.61 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2144: contig of 2144 bp in length
* 2145 2244: gap of unknown length
* 2245 4052: contig of 1808 bp in length
* 4053 4152: gap of unknown length
* 4153 6306: contig of 2154 bp in length
* 6307 6406: gap of unknown length
* 6407 8444: contig of 2038 bp in length
* 8445 8545: gap of unknown length
* 8545 11626: contig of 3082 bp in length
* 11627 11726: gap of unknown length
* 11727 15823: contig of 4097 bp in length
* 15824 15923: gap of unknown length
* 15924 20350: contig of 4427 bp in length
* 20351 20450: gap of unknown length
* 20451 25976: contig of 5526 bp in length
* 25977 26076: gap of unknown length
* 26077 31313: contig of 5237 bp in length
* 31314 31413: gap of unknown length
* 31414 38673: contig of 7260 bp in length
* 38674 38773: gap of unknown length
* 38774 45342: contig of 6569 bp in length
* 45343 45442: gap of unknown length
* 45443 53428: contig of 7986 bp in length
* 53429 53529: gap of unknown length
* 53529 60068: contig of 6540 bp in length
* 60069 60168: gap of unknown length
* 60169 68196: contig of 8028 bp in length
* 68197 68296: gap of unknown length
* 68297 76091: contig of 7795 bp in length
* 76092 76191: gap of unknown length
* 76192 87826: contig of 11635 bp in length
* 87827 87926: gap of unknown length
* 87927 99071: contig of 11145 bp in length
* 99072 99171: gap of unknown length
* 99172 15184: contig of 16013 bp in length

* 115185 115284: gap of unknown length
* 115285 132703: contig of 17419 bp in length
* 132704 132803: gap of unknown length
* 132804 154064: contig of 21261 bp in length
* 154065 154164: gap of unknown length
* 154165 173556: contig of 19392 bp in length.
Location/Qualifiers
1. .173556

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-507P19"
1. .2144
misc_feature
/note="assembly_name:Contig11"
2245. .4052
misc_feature
/note="assembly_name:Contig12"
4153. 6306
misc_feature
/note="assembly_name:Contig13"
6407. .8444
misc_feature
/note="assembly_name:Contig14"
8545. .11626
misc_feature
/note="assembly_name:Contig15"
11727. .15823
misc_feature
/note="assembly_name:Contig16"
15924. .20350
misc_feature
/note="assembly_name:Contig17"
20451. .25976
misc_feature
/note="assembly_name:Contig18"
26077. .31313
misc_feature
/note="assembly_name:Contig19"
31414. .38673
misc_feature
/note="assembly_name:Contig20"
38774. .45342
misc_feature
/note="assembly_name:Contig21"
45443. .53428
misc_feature
/note="assembly_name:Contig22"
53529. .60068
misc_feature
/note="assembly_name:Contig23"
60169. .68196
misc_feature
/note="assembly_name:Contig24"
68297. .76091
misc_feature
/note="assembly_name:Contig25"
vector_end:r7
vector_side:left"
76192. .87826
misc_feature
/note="assembly_name:Contig26"
87927. .99071
misc_feature
/note="assembly_name:Contig27"
99172. .115184
misc_feature
/note="assembly_name:Contig28"
115285. .132703
misc_feature
/note="assembly_name:Contig29"
clone_end:sp6
vector_side:left"
132804. .154064
misc_feature
/note="assembly_name:Contig30"
154165. .173556
misc_feature
/note="assembly_name:Contig31"
BASE COUNT 42375 a 43249 c 43427 g 42464 t 2041 others
ORIGIN

Query Match 41.2% Score 26.4; DB 59; Length 173556;
Best Local Similarity 69.2%; Pred. No. 40; Mismatches 16; Indels 0; Gaps 0;

Oy 12 tgcctagcctaagcgactagagctctgctgctctctctcgcgc 63
||||| ||||| | | | | | ||||| ||| ||||| ||||| | |
DB 140974 TGCTGTGGCTATGTCGGCGCAAGTTCCTTGCTTCTCTCTCTCC 140923

RESULT 10
AC015977/c

LOCUS AC015977 .207550 bp DNA HTG 17-AUG-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-527I18, WORKING DRAFT
AC015977
SEQUENCE: 23 unordered pieces.
AC015977
VERSION AC015977.7 GI:9838247
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 207550)
Waterston, R.H.
REFERENCE The sequence of Homo sapiens clone
AUTHORS unpublished
TITLE 2 (bases 1 to 207550)
REFERENCE Waterston, R.H.
AUTHORS Direct Submission
TITLE Submitted (17-NOV-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 17, 2000 this sequence version replaced gi:9211469.

51986 52085: gap of unknown length
* 52086 64048: contig of 11963 bp in length
* 64049 64148: gap of unknown length
* 64149 73711: contig of 9563 bp in length
* 73712 73811: gap of unknown length
* 73812 81475: contig of 7664 bp in length
* 81476 81575: gap of unknown length
* 81576 91689: contig of 10114 bp in length
* 91690 91789: gap of unknown length
* 91790 102726: contig of 10937 bp in length
* 102727 102826: gap of unknown length
* 102827 114269: contig of 11443 bp in length
* 114270 114369: gap of unknown length
* 114370 130207: contig of 15838 bp in length
* 130208 130307: gap of unknown length
* 130308 148001: contig of 17694 bp in length
* 148002 148101: gap of unknown length
* 148102 165221: contig of 18420 bp in length
* 165222 166379: gap of unknown length
* 166380 186479: gap of unknown length
* 186480 207550: contig of 21071 bp in length.

FEATURES

source

1. .207550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-527I18"
1. .2257
/note="assembly_name:Contig14"
2358. .4468
/note="assembly_name:Contig15"
4569. .7877
/note="assembly_name:Contig16"
7978. .11086
/note="assembly_name:Contig17
clone_end:SP6
vector_side:left"
1187. .14728
/note="assembly_name:Contig18"
14829. .19922
/note="assembly_name:Contig19"
20023. .25092
/note="assembly_name:Contig20"
25193. .30468
/note="assembly_name:Contig21"
30569. .34509
/note="assembly_name:Contig22"
34610. .40469
/note="assembly_name:Contig23"
40570. .46884
/note="assembly_name:Contig24"
46985. .51985
/note="assembly_name:Contig25"
52086. .64048
/note="assembly_name:Contig26"
64149. .73711
/note="assembly_name:Contig27"
73812. .81475
/note="assembly_name:Contig28"
81576. .91689
/note="assembly_name:Contig29"
91790. .102726
/note="assembly_name:Contig30"
102827. .114269
/note="assembly_name:Contig31"
114370. .130207
/note="assembly_name:Contig32"
130308. .148001
/note="assembly_name:Contig33"
148102. .165221
/note="assembly_name:Contig34"
166379. .186379

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Project Information
Center project name: H_NH0527I18
----- Summary Statistics -----
Sequencing vector: MJ3. 658
Sequencing vector: plasmid; 358
Chemistry: Dye-primer ET; 65% of reads
Chemistry: Dye-terminator Big Dye; 35% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 19353 bases at least Q40
Consensus quality: 198962 bases at least Q30
Consensus quality: 201706 bases at least Q20
Insert size: 201000; agarose-fp
Insert size: 207759; sum-of-contigs
Quality coverage: 4.48 in Q20 bases; sum-of-contigs
Quality coverage: 4.40 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2257: contig of 2257 bp in length
* 2258 2357: gap of unknown length
* 2358 4468: contig of 2111 bp in length
* 4469 4568: gap of unknown length
* 4569 7877: contig of 3309 bp in length
* 7878 7977: gap of unknown length
* 7978 11086: contig of 3109 bp in length
* 11087 11186: gap of unknown length
* 11187 14728: contig of 3542 bp in length
* 14729 14828: gap of unknown length
* 14829 19922: contig of 5094 bp in length
* 19923 20023: gap of unknown length
* 20023 25092: contig of 5070 bp in length
* 25093 25193: gap of unknown length
* 25193 30468: contig of 5276 bp in length
* 30469 30568: gap of unknown length
* 30569 34509: contig of 3941 bp in length
* 34510 34609: gap of unknown length
* 34610 40469: contig of 5860 bp in length
* 40470 40569: gap of unknown length
* 40570 46884: contig of 6315 bp in length
* 46885 46984: gap of unknown length
* 46985 51985: contig of 5001 bp in length

```
misc.feature /note="assembly_name:Contig35"
186480..207550
/note="assembly_name:Contig36
clone_end:T7
vector_side:left"
BASE COUNT 51842 a 51525 c 52063 g 49909 t 2211 others
ORIGIN
Query Match 41.2% Score 26.4; DB 40; Length 207550;
Best Local Similarity 69.2%; Pred. No. 41;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 12 tgcctagcactacgcacatcctctgcctcctctcctccgcgc 63
Db 143648 TGCTCTGGCTATGTGGGCACTTCCTCTCTCTCTCTCTCTCC 143597
RESULT 11
AC068757/c AC068757 215708 bp DNA HTG 06-SEP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-49A13, WORKING DRAFT SEQUENCE,
DEFINITION 41 unordered pieces.
ACCESSION AC068757.9 GI:9966609
VERSION AC068757.9
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 215708)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodoia,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durdin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,M., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Licharge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Stugang,R.,
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabhan,M.,
Wellington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 215708)
Worley,K.C.
Direct Submission
Submitted (09-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 5, 2000 this sequence version replaced gi:9438600.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBB
Center clone name: RP11-49A13
----- Summary Statistics
Assembly program: Phrap: bases 0.990329
Consensus quality: 173895 bases at least Q40
Consensus quality: 196384 bases at least Q30
Consensus quality: 205890 bases at least Q20
Estimated insert size: 191796; sum-of-contrigs estimation
```

```
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.3x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2183: contig of 21383 bp in length
* 21384 21483: gap of unknown length
* 21484 38359: contig of 16876 bp in length
* 38360 38459: gap of unknown length
* 38460 53950: contig of 15491 bp in length
* 53951 54050: gap of unknown length
* 54051 70489: contig of 16439 bp in length
* 70490 70590: gap of unknown length
* 70590 83686: contig of 13097 bp in length
* 83687 83786: gap of unknown length
* 83787 96752: contig of 12966 bp in length
* 96753 96852: gap of unknown length
* 96853 109306: contig of 12454 bp in length
* 109307 109406: gap of unknown length
* 109406 117664: contig of 8358 bp in length
* 117665 117864: gap of unknown length
* 117865 126080: contig of 8216 bp in length
* 126081 126180: gap of unknown length
* 126181 134443: contig of 8263 bp in length
* 134444 134544: gap of unknown length
* 134544 143313: contig of 8770 bp in length
* 143314 143413: gap of unknown length
* 143414 149730: contig of 6317 bp in length
* 149731 149830: gap of unknown length
* 149831 156223: contig of 6393 bp in length
* 156224 156323: gap of unknown length
* 156324 160771: contig of 4448 bp in length
* 160772 160871: gap of unknown length
* 160872 165489: contig of 4618 bp in length
* 165490 165589: gap of unknown length
* 165590 170132: contig of 4543 bp in length
* 170133 170232: gap of unknown length
* 170233 176439: contig of 6207 bp in length
* 176440 176539: gap of unknown length
* 176540 179371: contig of 2832 bp in length
* 179372 179471: gap of unknown length
* 179472 182074: contig of 2603 bp in length
* 182075 182174: gap of unknown length
* 182175 184961: contig of 2787 bp in length
* 184962 185061: gap of unknown length
* 185062 187133: contig of 2072 bp in length
* 187134 187233: gap of unknown length
* 187234 189001: contig of 1768 bp in length
* 189002 189101: gap of unknown length
* 189102 190458: contig of 1357 bp in length
* 190459 190558: gap of unknown length
* 190559 192237: contig of 1679 bp in length
* 192238 192337: gap of unknown length
* 192338 194506: contig of 2169 bp in length
* 194507 194606: gap of unknown length
* 194607 196046: contig of 1440 bp in length
* 196047 196146: gap of unknown length
* 196147 198227: contig of 2081 bp in length
* 198228 198327: gap of unknown length
* 198328 199813: contig of 1486 bp in length
* 199814 199913: gap of unknown length
* 199914 201119: contig of 1206 bp in length
* 201120 202411: gap of unknown length
* 202412 202511: contig of 1192 bp in length
* 202511: gap of unknown length
```


THIS PAGE BLANK (USPTO)


```

Db      2686 aaagcgtgttcgaattgctaacgcgtctaaagatatatgcacgcgttatgct 2738

RESULT 14
: US-08-466-662-2
: Sequence 2, Application US/084466662B
: Patent No. 6130059
: GENERAL INFORMATION:
: APPLICANT: Covacci, Antonello
: APPLICANT: Bugnoli, Massimo
: APPLICANT: Telford, John
: APPLICANT: Macchia, Giovanni
: APPLICANT: Rappuoli, Rino
: TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
: TITLE OF INVENTION: Diagnostics
: FILE REFERENCE: CHIR0057
: CURRENT APPLICATION NUMBER: US/08/466,662B
: CURRENT FILING DATE: 1995-06-06
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 2
: LENGTH: 3960
: TYPE: DNA
: ORGANISM: Helicobacter pylori
US-08-466-662-2

```

Query Match	32.8%	Score 21	DB 3	Length 3960
Best Local Similarity	62.3%	Pred. NO. 43		
Matches	33	Conservative	0	Mismatches 20; Indels 0; Gaps 0.
Qy	2	aagcccaatgtgctaaagcctaagcgaactagaagctcttcgtcgctgactctt	54	
Db	2606	aaagcgctcttgaatctgcttaacgcgtctctcaaaagatcatgcaacgcttatgct	2738	

RESULT 15
 US-08-200-232-1
 Sequence 1, Application US/08200232
 Patent No. 5721349
 GENERAL INFORMATION:
 APPLICANT: Cover, Timothy L.
 APPLICANT: Blaser, Martin J.
 TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. Pylori
 TITLE OF INVENTION: AND RELATED METHODS
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NEEDLE & ROSENBERG P.C.
 STREET: 127 Peachtree Street, Suite 1200
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/200,232
 FILING DATE:
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Spratt, Gwendolyn D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 2200.023
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ. ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4042 base pairs
 TYPE: nucleic acid

```

? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 101..3964
?
US-08-200-232-1

```

Query Match	32.8%	Score 21	DB 1	Length 4042
Best Local Similarity	62.3%	Pred. No.	43	
Matches 33	Conservative 0	Mismatches	20	Indels 0
				Gaps 0

Oy 2 aagcccaagtgtccttagcctaagcgacactagaatctcttgcctgcttgcttc 54
 || | || | | ||||| || | || | || | ||
Db 2742 AAGCGTCTTTGAATTGGCTAACGCGTCTTAAGATATTGACACGCTTATGCT 2794

Search completed: February 12, 2001, 21:25:16
Job time: 3402 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 19:56:04 ; Search time 1923.53 Seconds
(without alignments)
233.154 Million cell updates/sec

Title: US-09-600-602-1

Perfect score: 64
Sequence: 1 taagccagtggtgcttaagc.....gcttgctctctccgcctca 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estba.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estov1.*
68: em_estov2.*
69: em_estp11.*
70: em_estp12.*
71: em_estp13.*
72: em_estp14.*
73: em_estp15.*
74: em_estro1.*
75: em_estro2.*
76: em_estro3.*
77: em_estro4.*
78: em_estro5.*
79: em_estro6.*
80: em_estro7.*
81: em_estro8.*
82: em_estro9.*
83: em_estro10.*
84: em_estro11.*
85: em_estro12.*
86: em_estro13.*
87: gb_est41.*
88: gb_est42.*
89: gb_est43.*
90: gb_est44.*
91: gb_est45.*
92: gb_est46.*
93: gb_est47.*
94: gb_est48.*
95: gb_est49.*
96: gb_est50.*
97: gb_est51.*
98: gb_est52.*
99: gb_est53.*
100: gb_est54.*
101: gb_est55.*
102: gb_est56.*
103: gb_est57.*
104: gb_est67.*
105: gb_est68.*
106: gb_est69.*
107: gb_est70.*
108: gb_est71.*
109: gb_est72.*
110: gb_est73.*
111: gb_est74.*
112: em_esthum21.*
113: em_esthum22.*
114: em_esthum23.*
115: em_estom1.*
116: em_estom2.*

117: em_estp16:*
118: em_estp17:*
119: em_estp18:*
120: em_estp19:*
121: em_estp20:*
122: em_estp21:*
123: em_estp22:*
124: em_estp23:*
125: em_estp24:*
126: em_estp25:*
127: em_estp26:*
128: em_estp27:*
129: em_estp28:*
130: em_estp29:*
131: em_estp30:*
132: em_estp31:*
133: em_estp32:*
134: em_estp33:*
135: em_estp34:*
136: em_estp35:*
137: em_estp36:*
138: em_estp37:*
139: em_estp38:*
140: em_estp39:*
141: em_estp40:*
142: em_estp41:*
143: em_estp42:*
144: em_estp43:*
145: em_estp44:*
146: em_estp45:*
147: em_estp46:*
148: em_estp47:*
149: em_estp48:*
150: em_estp49:*
151: em_estp50:*
152: em_estp51:*
153: em_estp52:*
154: em_estp53:*
155: em_estp54:*
156: em_estp55:*
157: em_estp56:*
158: em_estp57:*
159: em_estp58:*
160: em_estp59:*
161: em_estp60:*
162: em_estp61:*
163: em_estp62:*
164: em_estp63:*
165: em_estp64:*
166: em_estp65:*
167: em_estp66:*
168: em_estp67:*
169: em_estp68:*
170: em_estp69:*
171: em_estp70:*
172: em_estp71:*
173: em_estp72:*
174: em_estp73:*
175: em_estp74:*
176: em_estp75:*
177: em_estp76:*
178: em_estp77:*
179: em_estp78:*
180: em_estp79:*
181: em_estp80:*
182: em_estp81:*
183: em_estp82:*
184: em_estp83:*
185: em_estp84:*
186: em_estp85:*
187: em_estp86:*
188: em_estp87:*
189: em_estp88:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	48.4	356	95	AW763066	AW763066 ur60d05.y
2	31	48.4	474	88	AW259340	AW259340 um92c02.y
3	31	48.4	485	147	W79969	W79969 me90d09.r1
4	31	48.4	533	20	A1425798	A1425798 me90d09.y
5	31	48.4	540	14	AA39633	AA39633 vz93h10.r
6	31	48.4	829	106	BE285743	BE285743 601096312
7	29.4	45.9	312	129	BB381871	BB381871 BB381871
8	27.4	42.8	746	107	BE415895	BE415895 MUC002.B0
9	26.2	40.9	221	172	AQ933849	AQ933849 RPCT-23-2
10	26.2	40.9	281	129	BB410001	BB410001 BB410001
11	26.2	40.9	296	130	BB410925	BB410925 BB410925
12	26.2	40.9	302	103	BB226614	BB226614 BB226614
13	25.8	40.3	465	182	AZ496732	AZ496732 1M033317
14	25.8	40.3	664	94	AW57473	AW57473 874001D08
15	25.8	40.3	764	175	A2118240	A2118240 RPCT-23-4
16	25.6	40.0	1052	29	AU050386	AU050386 AU050386
17	25.2	39.4	300	126	BB263403	BB263403 BB263403
18	25.2	39.4	328	132	BB497860	BB497860 BB497860
19	25	39.1	281	104	BE128709	BE128709 DEPA2459
20	25	39.1	283	139	C23376	C23376 C23376 Japa
21	25	39.1	378	40	AM126420	AM126420 614071C05
22	25	39.1	494	29	AU084524	AU084524 AU084524
23	25	39.1	515	21	A1503026	A1503026 vm75d11.x
24	25	39.1	526	157	AQ459497	AQ459497 HS-50B5_A
25	25	39.1	540	108	BE469225	BE469225 IPHdK0119
26	25	39.1	543	108	BE469051	BE469051 IPHdK0316
27	25	39.1	545	175	A2102103	A2102103 RPCT-23-1
28	25	39.1	548	181	A2401292	A2401292 IM0167L20
29	25	39.1	559	39	AW052990	AW052990 614077E04
30	25	39.1	580	88	AW259420	AW259420 up29d02.x
31	25	39.1	602	38	AW013436	AW013436 PC031KS W
32	25	39.1	626	39	AW053148	AW053148 614032H06
33	25	39.1	634	24	A1712055	A1712055 614004G06
34	25	39.1	860	97	AW983363	AW983363 HVSME9001
35	24.8	38.8	179	93	AW607959	AW607959 CM3-HT051
36	24.8	38.8	182	10	AA653836	AA653836 ns94h12.s
37	24.8	38.8	328	13	AA920859	AA920859 vy83f06.r
38	24.8	38.8	331	175	A2111044	A2111044 RPCT-23-4
39	24.8	38.8	458	18	A172462	A172462 uk06c02.y
40	24.8	38.8	524	174	A2079555	A2079555 RPCT-23-3
41	24.8	38.8	575	175	A2086580	A2086580 RPCT-23-2
42	24.8	38.8	685	173	AQ0985480	AQ0985480 RPCT-23-3
43	24.8	38.8	852	168	AQ745431	AQ745431 HS-2278_A
44	24.6	38.4	295	156	AQ390792	AQ390792 CITBI-EI-
45	24.6	38.4	325	184	B79194	B79194 CTT978SK-30

ALIGNMENTS

RESULT 1
LOCUS AW763066 356 bp mRNA
DEFINITION ur60d05.y1 NC1_CGAP_Mam3 Mus musculus cDNA clone IMAGE:3154665 5',
ACCSSION AW763066
VERSION AW763066.1 GI:7695001
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 356)
AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/tresources.shtml
MGI:1057421
Seq primer: -40RP from Gibco
High quality sequence stop: 332.
Location/Qualifiers
1. 356
/organism="Mus musculus"
/strain="129 - C57/B6 - FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3154665"
/clone_lib="NCI-CCGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMY-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."
BASE COUNT 93 a 57 c 89 g 117 t
ORIGIN
Query Match 48.4%; Score 31; DB 95; Length 356;
Best Local Similarity 72.7%; Pred. No. 1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY 5 cccagtgtgcttaagcactagaagcttctgctgcttcttc 59
|||||
Db 269 CCCAGTGTGCTGTACTTACTACAGCTCAGCTGCAGCGAGCTTCTCTCC 215
|||||
RESULT 2
LOCUS AM259340 474 bp mRNA EST 23-DEC-1999
DEFINITION um92c02.y1 Sugano mouse kidney mk1a Mus musculus cDNA clone
IMAGE:233226 5', mRNA sequence.
ACCESSION AM259340
VERSION AM259340.1 GI:6632321
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 474)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Sallier,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rittler
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1011430
Seq primer: custom primer used
High quality sequence stop: 466.
Location/Qualifiers
1. 474
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:233226"
/clone_lib="Sugano mouse kidney mk1a"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18S-FL3; Site.1: DraIII
(CACGCTG); Site.2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGACCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor (TGTGGCCCTACTGG), digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTTCGCTCTAAACCTCGG and 3' end
primer CGACCTCAGCTCGACCA."
BASE COUNT 133 a 70 c 112 g 156 t 3 others
ORIGIN
Query Match 48.4%; Score 31; DB 88; Length 474;
Best Local Similarity 72.7%; Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY 5 cccagtgtgcttaagcactagaagcttctgctgcttcttc 59
|||||
Db 321 CCCAGTGTGCTGTACTTACTACAGCTCAGCTGCAGCGAGCTTCTCTCC 267
|||||
RESULT 3
LOCUS W79969 485 bp mRNA EST 25-JUN-1996
DEFINITION me90d09.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:402833 5', mRNA sequence.
ACCESSION W79969
VERSION W79969.1 GI:1391071
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 485)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:246601
Seq primer: ETPRimer
High quality sequence stop: 346.
Location/Qualifiers
1. 485
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:402833"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCATCTGACAGTGGAGCGCGCGGAAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2); double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 139 a 94 c 109 g 143 t
ORIGIN

Query Match 48.4%; Score 31; DB 147; Length 485;
Best Local Similarly 72.7%; Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 5 cccagtgcttagcgaagcgcactgagctcttcgtcgtcttcctcc 59
|||||
Db 98 CCCAGTGTGCTGTAACCTTACTACGCTTCACGCTTCGACGCTTCTTC 44

RESULT 4
LOCUS A1425798 533 bp mRNA EST 15-MAR-2000
DEFINITION me90009.y1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:402833 5', mRNA sequence.
ACCESSION A1425798
VERSION A1425798.1 GI:4271729
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 533)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rittler
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)

TITLE JOURNAL
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESSEQUENCE of a previously sequenced mouse clone
MGI:246601
Seq primer: -40RP from Gibco

High quality sequence stop: 478
POLYA-NO.
Location/Qualifiers
1. 533
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:402833"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCATCTGACAGTGGAGCGCGCGGAAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2); double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 152 a 106 c 121 g 153 t 1 others
ORIGIN

Query Match 48.4%; Score 31; DB 20; Length 533;
Best Local Similarly 72.7%; Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 5 cccagtgcttagcgaagcgcactgagctcttcgtcgtcttcctcc 59
|||||
Db 112 CCCAGTGTGCTGTAACCTTACTACGCTTCACGCTTCGACGCTTCTTC 58

RESULT 5
LOCUS AA939633 540 bp mRNA EST 01-MAY-1998
DEFINITION v293h10.r1 Soares_thymus_2NDMT Mus musculus cDNA clone
IMAGE:1344835 5', mRNA sequence.
ACCESSION AA939633
VERSION AA939633.1 GI:3100410
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 540)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Schellenberg,K., Steptoe,M., Theising,B., Tan,F., Underwood,K., Moore,B.,
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

TITLE JOURNAL
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:693627

Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 496.
Location/Qualifiers

FEATURES 1. 540
Source /organism="Mus musculus"

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1344835"
/clone_1lb="Soares_thymus_2nbwT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer (5',
TGTTACCAATCTGAGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 159 a 97 c 116 g 168 t

ORIGIN

Query Match 48.4% Score 31; DB 14; Length 540;
Best Local Similarity 72.7% Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 5 cccagtgctgaagcgaacgacgtcttcgtcgtctctcttc 59
||||| 11111 11111 11111 11111 11111 11111
Db 179 CCCAGTGTGCTGTACTTACTTACTACGCTCTCAGCTGGCAGCTTCTCTCC 125

RESULT 6
BE285743/c 829 bp mRNA EST 13-JUL-2000
LOCUS 601096312F1 NCL.CGAP_Mam5 Mus musculus CDNA clone IMAGE:3490970 5',
DEFINITION mRNA sequence.
ACCESSION BE285743
VERSION BE285743.1 GI:9163574
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1AM8534 row: 9 column: 03
High quality sequence stop: 646.

FEATURES
source Location/Qualifiers
1. 829

/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:3490970"
/clone_1lb="NCL.CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: lothar Hennighausen/Robin Humphreys,

BASE COUNT 207 a 164 c 191 g 267 t
ORIGIN NIH"

Query Match 48.4% Score 31; DB 106; Length 829;
Best Local Similarity 72.7% Pred. No. 1.1;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 5 cccagtgctgaagcgaacgacgtcttcgtcgtctcttc 59
||||| 11111 11111 11111 11111 11111 11111
Db 261 CCCAGTGTGCTGTACTTACTTACTACGCTCTCAGCTGGCAGCTTCTCTCC 207

RESULT 7
BB381871/c 312 bp mRNA EST 13-JUL-2000
LOCUS BB381871 RIKEN full-length enriched, 0 day neonate cerebellum Mus
DEFINITION musculus CDNA clone C230016E14 3', mRNA sequence.
ACCESSION BB381871
VERSION BB381871.1 GI:9101601
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 312)
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Kusuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, T., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
TITLE JOURNAL
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp/
URL: <http://genome.rtc.riken.go.jp/>

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermostabilization of full length
cDNA and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, T., Ozawa, T., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
source Location/Qualifiers
1. 312
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230016E14"
/clone_1lb="RIKEN full-length enriched, 0 day neonate
cerebellum"

	/Issue_type="cerebellum"
	/dev_stage="0 day neonate"
	/lab_host="DH108"
	/note=Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'
	GAGGAGAGAAGCATCCACAGACTCTTTTTTTTATVN 3', cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second
	strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAATTTCGGATTAAATAAATTAATCCCCCCCCC 3'. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda
	FLC I "
BASE COUNT	90 a 72 c 66 g 84 t
ORIGIN	
Query Match	45.9%; Score 29.4; DB 129; Length 312;
Best Local Similarity	70.9%; Pred. No. 3.5;
Matches 39; Conservative	0; Mismatches 16; Indels 0; Gaps 0;
Db 101 CCCACTGTGCCTTAGCTTACTATACAGCTCGTGGAAGCTCTTCTCC 47	
QY 5 ccacagtcgtcttagcgaacgcacctagaacttctgctcgctctctcc 59	
BE415895/c LOCUS	
DEFINITION BE415895 746 bp mRNA EST 24-JUL-2000	
MUG002.B08R990520 ITEC MUG Wheat Spikulelet Library Triticum aestivum	
CDS clone MUG002.B08, mRNA sequence.	
BE415895	
BE415895.1 GI:9413741	
EST.	
bread wheat.	
Triticum aestivum	
Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.	
1 (bases 1 to 746)	
Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,	
S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,	
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Jouderier,P.,	
Lanegrade,P., Lazio,G.R., Lin,J.J., McGuire,P., Ogihara,H.,	
Pechioni,N., Quiset,C., Schuch,W., Selvaraj,G., Shariflou,M.,	
Sorrells,M., Warburton,M. and Wezel,G.	
International Triticale EST Cooperative (ITEC): Production of	
Expressed Sequence Tags for Species of the Triticeae	
Unpublished (2000)	
Contact: Ogihara Y	
Kihara Institute for Biological Research, Yokohama City University	
Maoka-cho 641-12, Tatsuka-ku, Yokohama 244-0813, JAPAN	
Tel.: 81 45 820 1903	
Fax: 81 45 820 1901	
Email: ogihara@yokohama-cu.ac.jp	
International Triticeae EST Cooperative (ITEC) http://wheat.pw.usda.gov/genome.	
Location/Qualifiers	
1..746	
/organism="Triticum aestivum"	
/cultivar="Norin 26"	
/db_xref="taxon:4565"	
/clone="MUG002.B08"	
/clone_lib="ITEC MUG Wheat Spikulelet Library"	
/tissue_type="young spikelets"	
/dev_stage="Feekes' scale 6-7"	
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:	

[illegible]

Query Match	40.9%	Score 26.2	DB 172	Length 221
Best Local Similarity	67.3%	Pred. NO. 41		
Matches 37	Conservative 0	Mismatches 18	Indels 0	Gaps 0
Db	16	AGCGGTTTCGTTTCCTCTCTCTGCTTCGCTCTCTTCGCTTCCTACACGCT	70	
Oy	8	agtgctcgaagcgaacgacactagagcttctgctcgtctcgtctcgcgt	62	
RESULT 10				
LOCUS	Bb410001/c			
DEFINITION	Bb410001 RIKEN full-length enriched, 7 days embryo Mus musculus			
ACCESSION	Bb410001			
VERSION	Bb410001.1			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryotic Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 281)			
AUTHORS	Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, N. and Hayashizaki, Y.			
TITLE	RIKEN mouse ESTs (Kono, H., et al.)			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Yoshihide Hayashizaki Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel: +81-298-36-9013 Fax: +81-298-36-9098 Email: genome-res@rtc.riken.go.jp/ URL: http://genome.rtc.riken.go.jp/ Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.			

FEATURES
SOURCE

Source

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C430016G23"
/clone_lib="RIKEN full-length enriched, 7 days embryo"
/dev_stage="7 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory
in

```

	Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5].
	GAGAGAAGAGATCCACAGACTCTTTTNTTTTTNNV N 3'}. cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently amplified for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5]
	GAGAGAAGATTCTCGAGTTAAATTAATTCGCCCCCCCCC 3'}. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from lambda ELC I."
BASE COUNT	80 a 60 c 67 g 74 t
ORIGIN	
Query Match	40.9%; Score 26.2; DB 129; Length 281;
Best Local Similarity	67.3%; Pred.No. 43;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;	
QY	cccaatgtgcttaagcactaaacgaactagaaccttcttgctgccttcctctctc 59
Dd	70 CCCAAGTGCTGTACCCTAACCTTCAACGCTCCTCAGCTGGAGGGGTTTCC 16

RESULT	11			
LOCUS	BB410925/c			
DEFINITION	BB410925	296 bp	mRNA	EST
ACCESSION	BB410925	16-JUL-2000		
VERSION	BB410925.1	CDNA clone C430020G13 3', mRNA sequence.		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			

REFERENCES

AUTHORS

TABLE
CONTENTS

COMMENT

C	22	36.4	15.0	14.6903	57	AC0686648	Homo sapi
C	23	35.4	14.8	15.4012	65	AC0709118	Homo sapi
C	24	35.8	14.8	16.2560	68	AL1358874	Homo sapi
C	25	35.8	14.8	17.0729	58	AC0589291	Homo sapi
C	26	35.8	14.8	17.9323	69	AL1391534	Homo sapi
C	27	35.8	14.8	19.6426	54	AC0320007	Homo sapi
C	28	35.8	14.8	22.1478	65	AC0320045	Homo sapi
C	29	35.6	14.7	19.6866	38	AC010742	Homo sapi
C	30	35.4	14.6	85.620	49	AC021298	Homo sapi
C	31	35.4	14.6	93.540	8	AB042235	Homo sapi
C	32	35.4	14.6	126.334	66	AL133403	Homo sapi
C	33	35.4	14.6	128.241	8	AB041340	Homo sapi
C	34	35.4	14.6	181.302	47	AL1355532	Homo sapi
C	35	35.4	14.6	300.000	38	AP002530	Homo sapi
C	36	35.2	14.5	193.227	38	AC010381	Homo sapi
C	37	35	14.5	154.256	58	AC073295	Homo sapi
C	38	35	14.5	170.356	11	AF259072	Homo sapi
C	39	35	14.5	201.833	84	CNS05722	Homo sapi
C	40	34.8	14.4	165.423	29	AC008111	Homo sapi
C	41	34.8	14.4	170.261	59	AC012315	Homo sapi
C	42	34.8	14.4	173.661	10	AC074107	Homo sapi
C	43	34.8	14.4	175.789	69	AP000425	Homo sapi
C	44	34.6	14.3	111.507	30	AC010001	Homo sapi
C	45	34.6	14.3	167.213	58	AC069549	Homo sapi

ALIGNMENTS

RESULT	1				
LOCUS	AR037064				
DEFINITION	Sequence from patent US 5801016.				PAT
ACCESSION	AR037064				
VERSION	AR037064.1	GI:5954920			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 173)				
TITLE	Morlocka,S. and Ueki,J.				
JOURNAL	DNA fragment, recombinant vector containing the same and method for expressing foreign genes using the same Patent: US 5801016-A 7-01-SEP-1998;				
FEATURES	Location/Qualifiers				
source	I..173				
BASE COUNT	21 a 48 c 47 g 57 t				
BIGIN					

```

Query Match      71.1%; Score 172; DB 81; length 173;
Best Local Similarity 100.0%; Pred. No. 2.7e-39;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

[illegible]

VERSION	AR037061.1	GI:5954917
KEYWORDS		
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified.	
AUTHORS	1 (bases 1 to 183)	
TITLE	Morioka,S. and Ueki,J.	
JOURNAL	DNA fragment, recombinant vector containing the same and method for	
FEATURES	expressing foreign genes using the same	
source	Patent: US 5801016-A 1 01-SEP-1998;	
	Location/Qualifiers	
	1..183	
BASE COUNT	23 a	51 c 51 g 58 t
ORIGIN	/organism="unknown"	

Query Match	71.1%;	Score 172;	DB 81;	Length 183;
Best Local Similarity	100.0%;	Pred. No. 2.7e-39;		
Matches 172; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY 71 taagcccaagtgtgctttagagcgaacgacacagagcttcctgacgcgtctgtctcttcgcg 130
 Db 7 TAACCCAGTGTGCTTAGGCTAAACGACATAGACCTTCTTGCGGCTGTCTTTCGCG 66
 QY 131 ctacagactgtctgtcgtcgtctgctagaaccctacactctgtcgtgagatgtcgtcg 190
 Db 67 CTCAGACTGTGCTTCTTGCTTCTTGCTTCCCTAGAACCCACTCTGTGCTGCAGATGTGCTG 126
 QY 191 ctctgctctctctccctcaagttcgatctgtatctgtgtgtggtgggggagcgag 242
 Db 127 CTTCGTCTCTCTTCTCTCAAGTTCGATCTGTATGTGTGTGTGGGGGGCGAG 178

AR005013	2799 bp	DNA	PAT	04-DEC-1998
Sequence	5	from patent US 5747327.		

[illegible]

Query Match	71.18;	Score 172;	DB 81;	Length 2799;
Best Local Similarity	100.08;	Pred. No. 3e-39;		
Matches 172;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

AR037063	AR037063	2799 bp	DNA	PAT	29-SEP-1999
LOCUS	Sequence	4 from patent US 5801016.			
DEFINITION	AR037063				
ACCESSION	AR037063.1	GI:5954919			
VERSION					
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 2799)				
TITLE	Morioka,S. and Ueki,J.				
JOURNAL	DNA fragment, recombinant vector containing the same and method for				
FEATURES	expressing foreign genes using the same				
SOURCE	Patent: US 5801016-A 4 01-SEP-1998;				
	Location/Qualifiers				
	1..2799				
	/organism="unknown"				
BASE COUNT	692 a	709 c	609 g	789 t	
ORIGIN					

	Query Match	71.1%	Score 172	DB 81	Length 2799;	
	Best Local Similarity	100.0%	Prod. No.	3e-39;		
	Matches	172;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
Oy	71	taagccagatgcttgaagctaagcgacataagacttccttgctgcgtttccttcctcgg	130			
Db	1667	TAAGCCAGATGCTTGAAGCTAAGCGCACTACAGCTTCCTGCTGCCTTTCTTCGG	1726			
Oy	131	cacagaatcgtctgcttgccttgcttcgcctagaaccctaacctctgacctgcagatgctgctg	190			
Db	1727	CTCAGATCTGCTTGCTTCTGCTTCGTCCGTAGAACCTACTCTGTCTCGTAGATGTCGCTG	1786			
Oy	191	cttgctcttccttcctcaagtccaattcagatctgatgtgtgtgtggtggggggccag	242			
Db	1787	CTTGCTTTCTCTCCTCAAGTTGCAATGTAATTGTGTGTGTGGGGGGCCAG	1838			

RESULT	5				
LOCUS	AB001920	5871 bp	DNA	PLN	14-APR-2000
DEFINITION	Oryza sativa DNA for phospholipase D, complete cds.				
ACCESSION	AB001920				
VERSION	AB001920.1	GI:1902902			
KEYWORDS	phospholipase D.				
SOURCE	Oryza sativa (cultivar:koshihikari) leaf DNA.				
ORGANISM	Oryza sativa				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.				
AUTHORS	1 (bases 1522 to 1655: 1839 to 1983; 2524 to 4420)				
TITLE	Ueki,J., Morioka,S., Komari,T. and Kumashiro,T. Purification and characterization of phospholipase D (PLD) from rice (Oryza sativa L.) and cloning of cDNA for PLD from rice and maize (Zea mays L.)				
JOURNAL	Plant Cell Physiol.	36 (5),	903-914	(1995)	
MEDLINE	96012933				
REFERENCE	2 (bases 1 to 5871)				
AUTHORS	Morioka,S., Ueki,J. and Komari,T.				
TITLE	Characterization of two distinctive genomic clones (Accession Nos. AB001919 and AB001920) for phospholipase D from rice (PGR97-076)				
JOURNAL	Plant Physiol.	114,	396	(1997)	
REFERENCE	3 (bases 1 to 5871)				
AUTHORS	Ueki,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-OCT-1995) to the DDBJ/EMBL/GenBank databases. Jun Ueki, Japan Tobacco Inc., Plant Breeding and Genetics Research Lab, 700 Higashihara, Iwata, Shizuoka 438-0802, Japan (E-mail:jun.ueki@pbgrl.jti.co.jp, Tel:81-538-32-7111, Fax:81-538-32-8700)				
FEATURES	Location/Qualifiers				
source	1..5871				
	/organism="Oryza sativa"				
	/cultivar="Koshihikari"				

mrna	1522..1665	1839..1983	2524..4420	4899..5702
exon	1522..1665	1838	2524..4420	4899..5702
intron	1666..1838	1839	4421..4898	
exon	1839	1983	4899..5702	
intron	1984..2523			
exon	2524..4420			
intron	4421..4898			
exon	4899..5702			
BASE COUNT	1488 a 1363 c 1355 g 1665 t			
ORIGIN	/db_xref="taxon:4530" /tissue_type="leaf" join(1522..1665,1839..1983,2524..4420,4899..5702) /citation=[1] 1522..1665 /number=1 1666..1838 /number=1 1839..1983 /number=2 join(1876..1983,2524..4420,4899..5332) /EC_number="3.1.4.4" /catalon=[1] /catalon=[1] /codon_start=1 /product="phospholipase D" /protein_id="BAAL9467.1" /db_xref="GI:1902903" /translation="MAQMLHGTLHATTFEASLSNPHRASGSPKIRKVEESIEDT VGKGVATIKVYSTIDLEKARVGRTRMTNEINERKSEFHIYAHMASNYIFPVKID NPGATINIGRAYLPVOLLNGEELDRMLDIDNNREVSGEKHKVLYQEPVSDRMN ARGRTNIGPVGVPYTFESQROGCKVYPLDYAHVVDNFKPLPLDAGNENRHEMED FDSLANSOHLIYITGMSVYTEITLVRSNRPKPGCVTLGELKKSGEYRMLVNW DFRSVALKRDGLMATHDEETENYPHGSNDVNCVLCPRNPDGSGIVSLISMFTFR HOKIYVYDHLPNGSSOORIVSYFVGLDCDGYDYQYHSLPFTLDSTHNDHPORP PNAIKSGGPREPMHDHSRLBGPIMADVLYNFEQRNRKOGKDLQLDLSDTII PPSVPMFGDEPRETNVOLFRIIDGGAIFGPDPTEEAKAGLVSGKDOIIDRSIDAA IHIIRAKNFIYIENOYFLGSSYAMKPEGIKPEDIGALHLPLKALVKSIEAGEE FYIVVYVPMPEGVPEGESVOAILDMQRTMEMMYITDTEMLQAKGTEANKDYLTFF CLGIREYKAGEYQEPEDPTDYSAOERREMIVYVHTKMLVDEVEYIIGSANTIN QRSMDGARDESIAMGYOPHLATROPARQIHFERMALWEHIGMLDYNQPRESLP CVQKVNRIARKYMDMYSDDLQDLPBGLHLSYPLGVASDGVYTELPKGEITFPDTRAR YLGASDYMPLPILTS" 1984..2523 /number=2 2524..4420 /number=3 4421..4898 /number=3 4899..5702 /number=4			

	Query Match	71.1%; Score 172; DB 6;	
	Best Local Similarity	100.0% ; Pred. No. 3e-39;	
Matches	172;	Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	71	taagcccaagtgtgctctaagaacgcacaatagaacttcctgatcggatccttctaccg	130
Db	1667	TAAGCCAGTGTGGCTTAGGCATAAAGGCACTAGAACCCTTGTCGTTCCATTCTCGC	1726
OY	131	cctcacattcgcttgcttgtccttgcttcgcttagaacacctcatctgtgctygcagttgcgtg	190
Db	1727	CCTGAATGTGGCTTGCTTCCTTGCTTCGTACGAACCTACTGTCCTCATGATGTGGCTG	1786
OY	191	cctgcgtctccctcccgaagtcgaatcgcattgctgtgtgtgagggggggccaa	242
Db	1787	CTTGCTCTTCTCTCTCTCACATTCGATCTGATTTGTGTGTGGGGGGGCGCAG	1838
RESULT	6		
LOCUS	166494		
DEFINITION	Sequence 7218 bp DNA	PART	28-Dec-1997
VERSION	166494		
KEYWORDS	166494.1 GI:2724471		
SOURCE	.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 7218)		
	Dorner,F., Schefflinger,F. and Falkner,F.Gunter,		


```
* 44143 44242: gap of 100 bp
* 44243 45375: contig of 1133 bp in length
* 45376 45475: gap of 100 bp
* 45476 48030: contig of 2555 bp in length
* 48031 48130: gap of 100 bp
* 48131 50762: contig of 2632 bp in length
* 50763 50862: gap of 100 bp
* 50863 54865: contig of 4003 bp in length
* 54866 54965: gap of 100 bp
* 54966 58786: contig of 3821 bp in length
* 58787 58886: gap of 100 bp
* 58887 62073: contig of 3193 bp in length
* 62080 62179: gap of 100 bp
* 62180 66768: contig of 4589 bp in length
* 66769 66868: gap of 100 bp
* 66869 70876: contig of 4008 bp in length
* 70877 70976: gap of 100 bp
* 70977 75996: contig of 5020 bp in length
* 75997 76096: gap of 100 bp
* 76097 80273: contig of 4177 bp in length
* 80274 80373: gap of 100 bp
* 80374 85679: contig of 5306 bp in length
* 85680 85779: gap of 100 bp
* 85780 91717: contig of 5938 bp in length
* 91718 91817: gap of 100 bp
* 91818 97046: contig of 5229 bp in length
* 97047 97146: gap of 100 bp
* 97147 101127: contig of 3981 bp in length
* 101128 101237: gap of 100 bp
* 101238 106788: contig of 5561 bp in length
* 106789 106888: gap of 100 bp
* 106889 111554: contig of 4666 bp in length
* 111555 111654: gap of 100 bp
* 111655 118598: contig of 6944 bp in length
* 118599 118698: gap of 100 bp
* 118699 124207: contig of 5509 bp in length
* 124208 124307: gap of 100 bp
* 124308 130324: contig of 6017 bp in length
* 130325 130424: gap of 100 bp
* 130425 138528: contig of 8104 bp in length
* 138529 138628: gap of 100 bp
* 138629 146991: contig of 8363 bp in length
* 146992 147091: gap of 100 bp
* 147092 153531: contig of 6440 bp in length
* 153532 153631: gap of 100 bp
* 153632 162631: contig of 9000 bp in length
* 162632 162731: gap of 100 bp
* 162732 174676: contig of 11945 bp in length
* 174677 174776: gap of 100 bp
* 174777 186910: contig of 12134 bp in length
* 186911 187010: gap of 100 bp
* 187011 201251: contig of 14241 bp in length.
*
FEATURES
Source
1..201251
Location/Qualifiers
1..201251
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/map="11"
/clone="RP23-136C7"
/clone_id="RP23-136C7"
1..1487
/misc_feature
/feature="assembly_fragment"
1588..2793
/misc_feature
/feature="assembly_fragment"
2894..4247
/misc_feature
/feature="assembly_fragment"
4348..5849
/misc_feature
/feature="assembly_fragment"
5950..7248
/misc_feature
/feature="assembly_fragment"
7349..8876
/misc_feature
/feature="assembly_fragment"
8977..10439
```

```
misc_feature
/feature="assembly_fragment"
10540..12080
misc_feature
/feature="assembly_fragment"
12181..14702
misc_feature
/feature="assembly_fragment"
14803..16733
misc_feature
/feature="assembly_fragment"
16834..18755
misc_feature
/feature="assembly_fragment"
18856..20786
misc_feature
/feature="assembly_fragment"
20887..23125
/feature="assembly_fragment"

Query Match 18.2% Score 44; DB 52; Length 201251;
Best Local Similarity 51.0%; Pred. No. 0.023;
Matches 104; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 36 ctcttgctgcgtgcttcttcctcgcgcacagatcccaagccagtgctagagcaagc 95
Db 42170 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 42111

QY 96 gcacagagcttctgctgcgtgcttcttcctcgcgcacagatcgcgtgctgcttct 155
Db 42110 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 42051

QY 156 cgcctagaccctactcctgctgctgcagtgctgcgtgctgcttcttcctcgcacagtgca 215
Db 42050 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 41991

QY 216 tctgattgtgtgtgtggtggggggcg 239
Db 41990 TGTGTGTGTGTGTGTGTGTGTGTG 41967

RESULT 8
AC026682 204806 bp DNA HTG 06-OCT-2000
LOCUS Mus musculus chromosome 11 clone RP23-201K2, WORKING DRAFT
DEFINITION
SEQUENCE 6 unordered pieces.
ACCESSION AC026682.9 GI:10645292
VERSION AC026682.9
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 204806)
AUTHORS Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okunou,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,T., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hoque,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louised,H., Lozano,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogun,M., Parish,B.,
Perez,L., Reltter,D., Say,J., Shen,X., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 204806)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 5, 2000 this sequence version replaced gi:9930744.
COMMENT Genome Center
```



```

----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 169222 bases at least Q40
Consensus quality: 170056 bases at least Q30
Consensus quality: 170433 bases at least Q20
Insert size: 170741; sum-of-contigs
Insert size: 171184; 25.5% error; agarose-fp
Quality coverage: 5.17% in Q20 bases; sum-of-contigs
coverage: 5.57% in Q20 bases; agarose-fp
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1      3832: contig of 3832 bp in length
*      3833 3932: gap of 100 bp
*      3933 165252: contig of 161320 bp in length
*      165253 165352: gap of 100 bp in length
*      165353 167736: contig of 2284 bp in length
*      167737 167836: gap of 100 bp
*      167837 171041: contig of 3205 bp in length.
Location/Qualifiers
1..171041
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-185C18"
/clone_11b="RPC11-1.1"
1..3832
/note="assembly_fragment:00817
fragment_chain:1
clone_end:SP6
vector_side:left"
3933..165252
/note="assembly_fragment:02691
fragment_chain:1"
165353..167736
/note="assembly_fragment:02301"
167837..171041
/note="assembly_fragment:02333"

BASE COUNT      50668 a 34334 c 35684 g 50052 t      303 others
ORIGIN

Query Match      16.3%; Score 39.4; DB 66; Length 171041:
Best Local Similarity 51.4%; Pred. No. 0.48;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0.

QY      1      gtaagccagtggtgcttaagcctaagcagcagctctctgctgctctctcc 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19991 GTAGAGCTAATAGGAGCTAGATGAGTGCAGTGGGATACAGATGTATTTCAAGTGCA 20050

QY      61      gtcagatcctaaagcccaagtgcttaagcctaagcgcacatagctctgctgtgc 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20051 GAGAAATAATGACAAAGGCGGGTGTGGCAATATATCTTAAAGACGTTACACCTCTTTC 20110

QY      121      ttctctccgcgtcaacatctgctgctgctgctgctgcgtctagaacactactgtgct 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20111 TACCTCTGCTTGCAATTTTGGCTTCTCTTTTCCTTTCCTTGCAACCTGTGTTATCT 20167

RESULT 12
AC079744      135697 bp      DNA      HNG      10-SEP-2000
LOCUS      AC079744
DEFINITION      Homo sapiens chromosome UNK clone CID-2326X17, *** SEQUENCING IN
PROGRESS ***, 31 unordered pieces.
ACCESSION      AC079744

```

VERSION	AC079744.1	GI:10047923
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
TITLE	1 (bases 1 to 135697)	
JOURNAL	Waterston,R.H.	
REFERENCE	2 (bases 1 to 135697)	
AUTHORS	Waterston,R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-SEP-2000) Genome Sequencing Center, Washington	
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
	MO 63108, USA	
COMMENT	----- Genome Center -----	
	Center: Washington University Genome Sequencing Center	
	Center code: WUGSC	
	Web site: http://genome.wustl.edu/gsc/index.shtml	
	----- Project Information -----	
	Center project name: H_MS23326K17	
	----- Summary Statistics -----	
	Sequencing vector: M13; 100%	
	Sequencing vector: plasmid; 0%	
	Chemistry: Dye-Primer ET; 100% of reads	
	Chemistry: Dye-terminator Big Dye; 0% of reads	
	Assembly program: Phrap; version 0.990319	
	Consensus quality: 116218 bases at least Q40	
	Consensus quality: 123738 bases at least Q30	
	Consensus quality: 126348 bases at least Q20	
	Insert size: 104000; agarose-fp	
	Insert size: 132052; sum-of-contigs	
	Quality coverage: 4.24 in Q20 bases; agarose-fp	
	Quality coverage: 3.44 in Q20 bases; sum-of-contigs	

	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 31 contigs. The true order of the pieces	
	* is not known and their order in this sequence record is	
	* arbitrary. Gaps between the contigs are represented as	
	* runs of N, but the exact sizes of the gaps are unknown.	
	* This record will be updated with the finished sequence	
	* as soon as it is available and the accession number will	
	* be preserved.	
	*-----	
	1	645: contig of 645 bp in length
*	646	745: gap of unknown length
*	746	1955: contig of 1210 bp in length
*	1956	2055: gap of unknown length
*	2056	3319: contig of 1264 bp in length
*	3320	3419: gap of unknown length
*	3420	4702: contig of 1283 bp in length
*	4703	4802: gap of unknown length
*	4803	6106: contig of 1304 bp in length
*	6107	6206: gap of unknown length
*	6207	7592: contig of 1386 bp in length
*	7593	7692: gap of unknown length
*	7693	9298: gap of 1606 bp in length
*	9299	9398: gap of unknown length
*	9399	10800: contig of 1402 bp in length
*	10801	10900: gap of unknown length
*	10901	12209: contig of 1309 bp in length
*	12210	12309: gap of unknown length
*	12310	13776: contig of 1467 bp in length
*	13777	13876: gap of unknown length
*	13877	15031: contig of 1155 bp in length
*	15032	15131: gap of unknown length
*	15132	16621: contig of 1490 bp in length
*	16622	16721: gap of unknown length
*	16722	18441: contig of 1720 bp in length
*	18442	18541: gap of unknown length
*	18542	20385: contig of 1844 bp in length
*	20386	20485: gap of unknown length

REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 144239)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
Summary Statistics
Center project name: H_RG333f24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:
Clone CTA-333f24 is from a release of the human BAC library CTB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBel0BAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTA-242D1, 200 bp overlap. The actual start of this clone is at base position 195 of CTA-333f24; actual end is at 145016 of CTA-333f24. This clone is part of an unanchored island, orientation is unknown.

The clone CTA-333f24 contains a transposable element, from bacterial DNA, starting at position 33810 and ending at 34586.

FEATURES

Source Location/Qualifiers
1..144239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="CTA-333f24"
/clone_1bp="CTB-HS-A"
/map="7q11.2-q21"
531..829
/rpt_family="Alu"
1072..1484
/rpt_family="MaLR"
1576..2103
/rpt_family="Retroviral"
3406..3596
/rpt_family="MIR"
4315..4481
/rpt_family="L2"
5683..5741
/rpt_family="L2"
5844..6210
repeat_region

/rpt_family="Retroviral"
7289..7376
/rpt_family="Achobo"
7384..7851
/rpt_family="MaLR"
7852..8374
/rpt_family="Achobo"
8373..8593
/rpt_family="Achobo"
8594..8727
/rpt_family="Alu"
8728..8898
/rpt_family="Achobo"
8988..9308
/rpt_family="Retroviral"
9543..9638
/rpt_family="MIR"
9658..9805
/rpt_family="Alu"
11635..11813
/rpt_family="MIR"
12320..12534
/rpt_family="MIR"
12751..13078
/rpt_family="MER2_type"
15243..15393
/rpt_family="L1"
16127..16486
/rpt_family="MaLR"
16487..18049
/rpt_family="MaLR"
18052..18390
/rpt_family="MaLR"
21329..22032
/rpt_family="L2"
28704..28821
/rpt_family="L1"
30594..30661
/rpt_family="Retroviral"
32037..32156
/rpt_family="L2"
33691..33809
/rpt_family="L1"
33828..34184
/rpt_family="MaLR"
34247..34513
/rpt_family="Alu"
34516..35404
/rpt_family="L1"
36874..37266
/rpt_family="MaLR"
37721..38010
/rpt_family="Alu"
39885..40072
/rpt_family="MIR"
40109..40407
/rpt_family="Alu"
41954..42551
/rpt_family="L1"
42639..42713
/rpt_family="L1"
43553..46528
/rpt_family="L1"
47407..47901
/rpt_family="MER21_g"
47929..48048
/rpt_family="L2"
50475..50560
/rpt_family="L2"
52156..52208
/rpt_family="L2"
52380..53628
/rpt_family="L1"
repeat_region

[illegible][illegible]

```
misc_feature      /note="assembly_fragment"
26413. .28720
/note="assembly_fragment"
misc_feature      28821. .31342
/note="assembly_fragment"
misc_feature      31443. .34159
/note="assembly_fragment"
misc_feature      34260. 35873
/note="assembly_fragment"
misc_feature      35974. .38802
/note="assembly_fragment"
misc_feature      38903. .41723
/note="assembly_fragment"
misc_feature      41824. .45044
/note="assembly_fragment"
misc_feature      45145. .48071
/note="assembly_fragment"
misc_feature      48172. .52438
/note="assembly_fragment"
misc_feature      52539. .55979
/note="assembly_fragment"
misc_feature      56080. 60454
/note="assembly_fragment"
misc_feature      60355. .63552
/note="assembly_fragment"
misc_feature      65453. .69953
/note="assembly_fragment"
misc_feature      70054. .77023
/note="assembly_fragment"
misc_feature      77124. .86478
/note="assembly_fragment"
misc_feature      86579. .98029
/note="assembly_fragment"
```

	Query Match	15.7%	Score 38;	DB 58;	Length 116534;	
	Best Local Similarity	49.0%	Pred. No. 1,2;	Mismatches	105;	Gaps
	Matches 101; Conservative	0;				
Oy	21	ctaaagcgaactagaagcttcgtcgctgcgtctctctccgacccaagtccaagaccgaagc 80				
Db	115546	CTTAGTTCCTCCTTTTCCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 115605				
Oy	81	gtgccttaggctaagcgaactagaagcttcgtcgctgcgtctctctccgacccaagctg 140				
Db	115606	TGTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 115665				
Oy	141	cttgcttgcgttcgtcgtcgtatagaacctactcgtcgtcgagtgctgcgtcttcctc 200				
Db	115666	CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 115725				
Oy	201	cttcctcaagtcgatcgtatgctgt 226				
Db	115726	CCTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 115751				
RESULT 15	AC009602/c					
LOCUS	Leishmania major chromosome 35 clone L2259 strain Friedlin,					
DEFINITION	complete sequence.					
ACCESSION	AC009602					
VERSION	AC009602.6 GI:8737222					
KEYWORDS	HMG.					
SOURCE	Leishmania major.					
ORGANISM	Leishmania major Eukaryote; Euglenozoa; Kinetoplastida; Trypanosomatidae; ,					
REFERENCE	1 (bases 1 to 37782)					
AUTHORS	Wyller,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.					
TITLE	Cawtha,J., Maisolini,F., Sunkin,S. and Stuart,K.D.					
JOURNAL	Direct Submission Submitted (28-AUG-1999) Seattle Biomedical Research Institution,					
	Nickerson Street, Seattle, WA 98109-1651, USA					

REFERENCE 2 (bases 1 to 37782)
AUTHORS Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Stuart,K. and Ivens,A.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
3 (bases 1 to 37782)
AUTHORS Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Stuart,K., Ivens,A. and Worthy,E.A.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
On Jun 26, 2000 this sequence version replaced gi:7651790.
NOTE: This sequence is still preliminary and may contain some
errors, such as substitutions and frameshifts. Please regard all
CDS sequences with some degree of caution. A finished version
should be available within a few weeks.
FEATURES
source
1..37782
Location/Qualifiers
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="L2259"
complement(<1..4125)
/gene="L2259.1"
/note="orf35.0650; predicted using Testcode and
Codonusage. ORF begins at 4125, Testcode and Codonusage
predict that 5th ATG at 3996 is used as start codon"
/gene="L2259.1"
complement(<1..3993)
/note="Blastp similarity to MAP kinases in a number of
different organisms; Blastc similarity to PRP8 protein
homologs and splicing factors in a number of different
organisms;Eukaryotic protein kinase domain, TPR domain and
HECT-domain (ubiquitin-transferase) identified by Pfam;
identified as belonging to the family of Serine/threonine
protein kinases by COGS; HECT domain profile, Protein
kinase domain profile, TPR repeat region, circular profile
and TPR repeat profile identified by PROSITE;
Serine/threonine protein kinases active-site signature
pattern identified by PROSITE; Tyrosine protein kinases
specific active-site signature identified by Gribskov;
dileucine motif and coiled-coil region identified by
PSORT."
/codon_start=1
/product="L2259.1"
/protein_id="AAG22839.1"
/db_xref="GI:10765025"
/translation="MPLSLQAMGFSVLVSRGPGVVSAGPTRGAEASPMAPDAOS
ASPGSVSGGVPMWVIOREQRIPLQFSATGSPNDGTRLREGEAKAGRYHEAI
RYTQAIEVDPSSEFLYTNRSFAVENIKFEKSSADAARAVEINANFPGHYLGAO
MSLNDGHAMESLRKAMALAPSENKEAIVRAAKCESKMARAVPPLMSGGSSVP
YTEGVSRSNPTAPRPTDAOPALGSTSDPCEOAETRIHAATLRDAADATYRR
ADDAIVECKSRSDKISLSTSSLSLAKEDKQSOQLRAVADDSNATYHAKRC
DNDATLSKLMDTRAQGMALQDKTKLAKEEQNFSEFGASNALSINVTGEGSDLT
AAAPVAVASPPASAPVPPNEAPADAGAPTSLSRASCSIPGSSPTAAVAPHOPL
TORAGHNHTOSTTAGALDQLLRRIEINERVAQSCFATDAAGAFYLSLAE
AAMKEVRSLLAEAVELDTQLETHAROVAOSTTKELSDVRLADOMSSIVIEVKD
EAFVKTMRGGDGLLEEAKLBOORLIHERORILOLAEIMFKRIDPEPKTICIOY
KRLQRIALIHKKQAVQTRIMELVEQDIPELAKMSKANGSRRLRYKSGSLMONLS
SPFOVLTLSSTVNSKVYHALRGEHVAVKEISIDDAARRRREFRNVTVTCNPNLI
IRIKVFGDPAPAYLILPYRHGSLRALSKQPMQSVAVODMLRQLAGVAYLHGIG
IYHGDLPNSNVLADGPRVISPFGIAKDHALGVADMTLTNTVTNSGYSIVGTI
OYAPBQLCEGTSHAKSTGMSDMALGFTMLEVLAENAFPHDAISGKPSLILPE
AKRIDPAAVAGDEKLAEITIASLVADPRANRATADLLAHPEFSTSSMNSNONS
AAKSDERIDAVRSYITAVRSHQVLYSVSRNHVESVRDIFORLDDDLSPITAV
VQGEAGIDEGALITELMLEYQQLLVKALKVCASEEAGASPPASIVSSITAA
ASASAPNEGHSIVLESTVYLPAPDADGIADLFLKGLVKLSITERKQIPLOQS
SVKLPSGAEPSEFVLEEDPSIATSLRLRLSLADLADAGIDFETPHFEKKAQD
GYYTDSPLTPESVDYISLRKFPDLVERRPRLPAMRGFCPEPDLHHKILSPSD
LLILGGLHVASAOYIVDALEFGQFGRNSNTPKYLKEVLLDSQNNLRFLDLCTATA
AVPSGAMKTI"

gene
complement(6308..7390)
/gene="L2259.2"
/note="orf35.0651; predicted using Glimmer, Testcode and
Codonusage"
CDS
complement(6308..7390)
/gene="L2259.2"
/note="Blastp similarity to mitochondrial carrier proteins
and hypothetical mitochondrial proteins from a number of
different organisms; mitochondrial carrier protein motif
identified by Pfam; dileucine motif identified by PSORT;
Mitochondrial energy transfer proteins signature
identified by both PROSITE and Gribskov."
/codon_start=1
/product="L2259.2"
/protein_id="AAG22840.1"
/db_xref="GI:10765026"
/translation="MEVCSALAGIFARLLCHPLDIKTIPTFGAGDHFAAPTAAS
SRPAPTAFLOSARSIMAREGVAGYRRVGAALAGSAGVALISTYDWCSTANKAF
GDAAGAAAGMAGTGESPVGYMNAITQLPRAAATPSSVRFVCGLAETVSCVW
VPIDVAKERLQSQPSPSLKGRYTSLDALRIIVNECARGLYKASTLSFGPFSVY
FVVEYFTNMLAGLYAASAPTCGRNCGSKERSLATFVALGACAGQWVASIL
TNPLELVKTRICQRAVILHRDGGASTPALFSVHRGLREGIALAKKEGIMALKVC
GSRIAYTAPNALTMGEFFELKTTLM"
complement(15577..17967)
/gene="L2259.3"
/note="orf35.0652; predicted using Glimmer, Testcode and
Codonusage"
CDS
complement(15577..17967)
/gene="L2259.3"
/note="nuclear localization predicted by PSORT;
Glutamine-rich, glycine-rich and Proline-rich regions
identified by PROSITE; 7 transmembrane receptor (Rhodopsin
family), Eukaryotic aspartyl protease, Bowman-Birk serine
protease inhibitor family,
Glutamate/leucine/phenylalanine/Valine dehydrogenase,
kinesin motor domain, Neurotransmitter-gated ion channel,
Orcidine 5'-phosphate decarboxylases, Oxidoreductases,
nucleogense component 1 and other families,
phosphoglycerate kinases, Site-specific recombinases and
Regulator of G protein signalling domains identified by
Pfam; coiled-coil region identified by PSORT."
/codon_start=1
/product="L2259.3"
/protein_id="AAG22841.1"
/db_xref="GI:10765027"
/translation="MPSVRYTWYREELRIATLASAMDGPVLKRPDYIEAVPPVT
PKRRAESNFKGSVFTDNSTTTTINIGPNCNGSGGCGNSGSSNNPGR
SSTPYVSGSGGGRNRRGGGGGNGGDDSSNSVNSQSTDRAPRBEGRPNNGR
NOETEEGLEYELROSALOKKRVAVRSHFESAQPTATSAHPATTQPSPPQASTLLS
ENDTEERLLAGITIVDDAPKAEVRSRHFSAQPTATSAHPATTQPSPPQASTLLS
FGAPASSIPATTTGTPQASVLPPTPANSCTQGYARQWSMKSDSNLMSTAPMASALK
OSLOHSILPADASPSPVIRKPPQASQSQSPSCPASTSHQATSMCAANAANSATKA
PASAGSGGVAPRASPAPRPOSVSGPAGHSGSTNTSHQATSMCAANAANSATKA
PHTPPTQTHNHSQSPSRAPRTATIKARRSGSGTVPFGANSADLELLKKGATMI
SSAPRPPSTASATAPKTOPITIALESMLMQRRQGVTKGMSLVPPSPQFO
QPPRPPQARLTPANSKTPRPNANSPTSPASSPOULPASKMKNPMGMLQPKPQFO
QOQPLRPPRPPMOPRASHQSQPOPOPOPOLOPOQOIOPOQOITPMAMPARPPQAP
QNSNSPMHNTPLDRLAKMRPOQAPRMQONCLOPMQONNAGQGVFVDAPOALQRYMT
GQPMRPRRDQGVFMQDHPMPQPOQAPRQOQAPRPPNPGTMAQFLFQOQOQQRN"
complement(25479..27959)
/gene="L2259.4"
/note="orf35.0653; predicted using Glimmer, Testcode and
Codonusage. ORF begins at 27959, Testcode and Codonusage
predict that 3rd ATG at 27629 is used as start codon"
complement(25479..27629)
/gene="L2259.4"
/note="nuclear localization predicted by PSORT."
/codon_start=1
/product="L2259.4"
/protein_id="AAG22842.1"
/db_xref="GI:10765028"
/translation="MAHANGSGSLGMSERDMRLHSRYTPNRVSLPEFKSVSGTY
HNYLAKDLHKIMQELRLLYEHGELREASVYPLNGKKQKELSRRAATAVRLIDRSGDG

IALPARRRSTSLATGALGROGOSTPPTJSGSRMAADSLSTOPQAPAT
 STLSISPRRSRVPPLTIDHIOEGOSTIASHSLMTAPOTYLPSTPSASQSPPT
 BARBAAPRGSCOTPHASTASLSOTLSRPTSTSRANRSTYAHNAATSYAVVA
 PSTASAPVAPRSLTITLSTSSNELAQLNEASAPPPFLISYPRFROTBRESSMLRE
 IPVOYVGVANSRQLRTYMLPRLPRTPAKMLVAVYVNCQCMTWKRSWMLRRE
 EVASYTYPLDITLSSNTDAQIVONDHINKRQTPALLAAVALPLEEYETEMWLSSR
 LGCNAAVQARQMLSSSSNANALATSSLSLSNNALATQRYAATMDDEDDCLTAVD
 DPTVYTCQPLSQLOLTPVYVSGSCITLQCVDNALATVSNKSNAYNCALCAVDRPCD
 VRVOTYLMATSLCFESSYDAVPLTILRLAASRSPKEETVYHMHSPRTCEJEDACV
 EASDEENCRSTSSGVSPAAVSAVSSASMTAPACGSEANISVAPNGARLIPAPASSHRG
 EAMRECHGRKSRGSRNDITPTEPRRTADDTIEL*
 complement(31724, -32869)

genes

```

/note="orf35.0654; L2259.5; predicted using Glimmer,
Testcode and CodonUsage"
complement(31724..32869)

```

CDS

/gene="ЛРКАСНЗ"

note="Blasf/Blasf/Thlasn homology to cAMP-dependent protein kinase catalytic subunits from a number of different organisms; cytoplasmic localization identified by PSORT; Protein kinase C terminal and eukaryotic protein kinase domains identified by Pfam; identified as belonging to the family of Serine/threonine protein kinases by COG; Bipartite nuclear localization signal and Protein kinase domain profile identified by PROSITE; Protein kinases ATP-binding region signature and Serine/threonine protein kinases active-site signature identified by PROSITE."

```
/product="putative protein kinase A catalytic subunit"  
/protein_id="AAG22843.1"
```

```
/protein_id="AAG22843.1"
```

```
/db_xref="GI:10765029"
```

/LrnsiSLkLpVtMlYsXpREvNGTfHlLgTGSrNvRGAPtITAnSdSAMP
 STWStSLkLpVtMlYsXpREvNGTfHlLgTGSrNvRGAPtITAnSdSAMP
 LKREIILKMOOHOVAENGILMEILCHPITIYNMKSPODEKkYVtLEFWGGMPLH
 LRTfRGPRNDYkTfAEtLYSLAfEILStSLdIIRtOLKfEPRtLNDLKGvKtDfGfAK
 KVPDRITTLGCTfAEtLYSLAfEILStSLdIIRtOLKfEPRtLNDLKGvKtDfGfAK
 ELtLAGRtKfEPNMEdGARULvKGLQDTHTIRtGLtKdGPADvANdHEtFPGAMMDL
 YARvYPIPIvKtSPDSrNEFKvTSDvDTPALtSAQAELKfG

Query Match	15.58;	Score 37.4;	DB 31;	Length 37782;
-------------	--------	-------------	--------	---------------

Best Local Similarity 50.98; Pred. No. 1.7;

Matches	89;	Conservative	0;	Mismatches	86;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Q7 49 tgcctcctccgctcagatcactaacgccagttgcttagtcgaacgcacctagaagcttc 108
||||| || | |||| | |||| | |||| |
||||| |||| | |||| | |||| | |||| |

Db 22975 TGCCTTTTTCGTTGTTGAGCTTCGAGCAGTTAGTCTGCCCTGCATTGACCGTC 22916

QY 109 ttgctcgcttgcctctctccgctcagatctgcttgcttgcttgctagaacctta 168

Db 22915 CTCCTTCATCTGCTCCTCTCTCCTCTTCTATTCTCTGTTGCCCCCCTCCCTCCCTC 22856

QY 169 ctctgtgctgcgagtgtcgctgcttctcctcctcaagttcgaatctgattg 223

Db 22855 CTCTTCCCTCCCTCTCTACCTCCTCTCCCCCTTCCTCCCTCTCTCTCTCCGTG 22801

THIS PAGE BLANK (USPTO)

ID	Q86785 standard; DNA: 2799 BP.
XX	
AC	Q86785;
DT	04-MAR-1996 (first entry)
XX	
DE	DNA encoding Phospholipase D.
XX	
KW	phospholipase D; measurement; reagent; phospholipid level; ss.
XX	
OS	Zea mays.
XX	
FH	key
FT	exon
FT	location/Qualifiers
FT	1876..1983
FT	/*tag= a
FT	1984..2523
FT	/*tag= b
FT	2524..2799
FT	/*tag= c
FT	/note= "partial exon 2"
XX	
PN	M0509234-A1.
XX	
PD	06-APR-1995.
XX	
PF	30-SEP-1994; 94MO-JP01627.
XX	
PR	30-SEP-1993; 93JP-0267884.
XX	
PA	(NISB) JAPAN TOBACCO INC.
XX	
PI	Morioka S, Ueki J;
XX	
DR	WPI: 1995-147433/19.
DR	P-PsDB: R72799.
XX	
PT	Cloned DNA coding plant derived phospholipase D - controls
PT	expression of plant derived PLD gene
XX	
PS	Claim 14; Page 33-35; 41pp: Japanese.
XX	
CC	The DNA contains sequences necessary for the expression of a plant
CC	derived phospholipase D (PLD). The PLD is useful for measuring
CC	phospholipid levels and for producing derivs. by e.g. base exchange
CC	reactions.
XX	
SQ	Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other:
<hr/>	
Query Match	71.1%; Score 172; DB 16; Length 2799;
Best Local Similarity	100.0%; Pred.No. 5,7e-42;
Matches 172; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
<hr/>	
QY	71 taagcccaagtgtgcttaaggctaagcgacataagactcttgcgtcgttcctctccg 130
DG	1667 taagcccaagtgtgcttaaggctaagcgacataagactcttgcgtcgttcctctccg 1726
<hr/>	
QY	131 ctgaagatcgttgcttgcttgcttgcttgctagaacctactctgtctcgaattcgcgtg 190
DG	1727 ctcaagatcgttgcttgcttgcttgcttgctagaacctactctgtctcgaattcgcgtg 1786
<hr/>	
QY	191 ctctgcttctctctccaagttcgaatcgtatgtgtgtgtggggggcgacaa 242
DG	1787 ctctgcttctctctccaagttcgaatcgtatgtgtgtgtggggggcgacaa 1838
<hr/>	
RESULT	5
ID	T42854
XX	T42854 standard; cDNA to mRNA; 2799 BP.
AC	T42854;
XX	
DT	16-JUN-1997 (first entry)

	Phospholipase D gene sequence fragment.
XX	Phospholipase D; rice; promoter; ss.
XW	Oryza sativa.
OS	
XX	
FH	Key Location/Qualifiers
FT	exon 1876..1983
FT	/tag= a
FT	/number= 1
FT	/note= "encodes residues 1 to 36 of W06134"
FT	intron 1984..2523
FT	/tag= b
FT	/number= 1
FT	exon 2524..2799
FT	/tag= c
FT	/number= 2
FT	/note= "encodes residues 37 to 128 of W06134"
XX	
PN	W09630510-AI.
PD	03-OCT-1996.
XX	
PF	28-MAR-1996;
PR	96WO-JP00812.
XX	
PR	29-MAR-1995;
XX	95JP-0096126.
PA	(NISB) JAPAN TOBACCO INC.
XX	
PI	Morioka S., Ueki J;
DR	WPJ: 1996-455357/45.
DR	P-PADB; W06134.
PT	Promoter DNA sequence derived from rice - used to increase expression of foreign genes in transformed hosts
PS	Disclosure; Page 20-22; 29pp; Japanese.
CC	This sequence represents a fragment of the coding sequence of the rice phospholipase D gene (PLD). The promoter for the PLD gene was isolated using the primers shown in T42857 and T42588. The promoters (see T42851 and T42852) are efficient promoters for greatly increasing the expression of foreign genes in transformant rice and other plants.
SQ	Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;
Query Match	71.1%; Score 172; DB 17; Length 2799; Best Local Similarity 100.0%; Pred. No. 5.7e-42;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Dy	71 taagccacagtgtcttaaggctaagcgacaataagaacttcttgctgcgcttgtccttcgccg 130 Db 1667 taagccacagtgtgacctaaagcgcaatgaagctcttgctgcgcttgtccttcgccg 1726 OY 131 ctcacagattgcttgcttgcttgcttcggctagaaccctactctgtgctgcgagttgcgtq 190 Db 1727 ctcacagattgcttgcttgcttcggctagaaccctactctgtgctgcgagttgcgtq 1786 OY 191 ctctgcttcctctccaagtcgcatctgatctgtgtgtgtggggggcgcgacg 242 Db 1787 ctctgcttcctctccaagtcgcatctgatctgtgtgtgtggggggcgcgacg 1838
RESULT 6	A10594/C
ID	A10594 standard; DNA; 10732 BP.
XX	A10594;
AC	
XX	29-JUN-2000 (first entry)
T	

xx	DE	Gene encoding a subunit of cellulose synthase.
xx	KM	Cellulose synthase; cellulose production; increase yield; ds.
xx	OS	Vigna angularis.
xx	PN	JP2000060568-A.
xx	PD	29-FEB-2000.
xx	PF	26-AUG-1998; 98JP-0239998.
xx	PR	26-AUG-1998; 98JP-0239998.
xx	PA	(MIZU) MIZUNO K.
xx	PA	(OJIP) OJI PAPER CO.
xx	DR	WPI; 2000-342371/30.
xx	DR	P-PsDB; Y85179.
xx	PT	A gene encoding a cellulose synthetic equipment - for the improvement
xx	PT	In the amount of cellulose synthesised in a plant body
xx	PS	Claim 2; Page 14-21; 32pp; Japanese.
xx	CC	This sequence represents a gene encoding a subunit of the cellulose
xx	CC	synthase complex of Vigna angularis. The invention relates to subunits of
xx	CC	cellulose synthetic equipment, that can be used to increase the amount of
xx	CC	cellulose synthesised by a plant. The proteins and genes encoding them
xx	CC	can also be used to improve the properties of the cellulose being
xx	CC	produced by a plant.
xx	SQ	Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;
		Query Match 14.7%; Score 35.6; DB 21; Length 10732;
		Best local Similarity 15.9%; Pred. No. 0.2;
		Matches 36; Conservative 99; Mismatches 92; Indels 0; Gaps 0;
OY	15	cttaagcctaagcgcaactaagctctctgctgctgctctctcgcgcagactcctaag 74
DB	9668	BTTSKRYSTTYTSVSMCYDATTMSRCTBYSTBRTBSRCAKCTBDSTSTANSTSTYSTRK 9609
OY	75	cccaagtgctctaagcgaactaagcgaactctctgctgctgctctctcgcgcga 134
DB	9608	TBYSRSRNGYSVCSRSRNCYCYTDSDSTCYSTTYTAVSCSTTSRNGSYDASRST 9549
OY	135	gactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctc 194
DB	9548	YSRCTTSTYSYSTTYSDDSTTTBNSYSSDSCITBYSRSDSCRSRSDSTCNCYSCSDS 9489
OY	195	gctctcctcctcgaagtcgcatcgtatgltgltg999999gcga 241
DB	9488	KISTTYDACCTYDAKTBCTYTSNDNCNSTSRCTNYSYSCRTSR 9442
RESULT	7	
ID	X13856	
XX	X13856	standard; DNA: 796 BP.
XX	AC	X13856;
XX	DT	19-MAR-1999 (first entry)
XX	DE	Enterococcus faecalis genome contig SEQ ID NO:919.
XX	KM	Enterococcus faecalis; contig; detection; Enterococcal infection;
XX	KW	vaccine; attenuation; computer readable medium; ds.
XX	OS	Enterococcus faecalis.
XX	PN	WO9850555-A2

XX 12-NOV-1998.
PD
XX
XX 04-MAY-1998; 98WO-US08985.
XX
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046555.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Dillon PJ, Kunsch CA;
PI
DR WPI; 1999-045171/04.
XX
XX New isolated *Enterococcus faecalis* polynucleotides and polypeptides
PT - used to develop products for the detection of *Enterococcus* and for
PT use in vaccines for prevention or attenuation of *Enterococcus*
PT infection.
XX
XX Claim 1; Page 2047-2048; 2084pp; English.
PS
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the *Enterococcus faecalis* genome
CC X12938 to X12919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the *Enterococcus faecalis* genome with
CC commercial importance. The products can be used to detect the presence
CC of *Enterococcus faecalis* in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 796 BP; 113 A; 170 C; 163 G; 346 T; 4 other;
XX

Query Match	14.18	Score 34.2	DB 20	Length 796
Best Local Similarity	51.78	Pred. No. 0.25	Mismatches 73	Indels 0
Matches 78	Conservative	0	Mismatches	73
QY	63	tcagatccctaacgcccagtgctgcttagagctaaagcgacataagagctctctgctgcgtgctt	1222	
Db	368	tcacgcgcgaagcagtgatttgcttctctctgctgctgctactctctgctgtgctgcgtgctt	427	
QY	123	ctcttcgcgtcagatctgcgtctgcttgccttgcctgcgcagaacctactctgtgctgcgag	1822	
Db	428	gactcgcgcacagcttgcgtctgcttcaatgcttcgcgcgtatagacttccaagctgcgtcgcg	487	
QY	183	tgctgcgtcgtcgtctccctccctcccaagtct	213	
Db	488	tgtagctcttcttcttgctggtctcccgcttctc	518	
RESULT	8			
TS5666				
ID	TS5666	standard; DNA; 2903 BP.		
XX	XX			
XX	XX	TS5666;		
XX	DT			
XX	07-FEB-1997	(first entry)		
XX	XX			
XX	DE	Partial yeast gene, Lg-FLO1, involved in flocculation.		
XX	KW	Flocculation; aggregation; floating; yeast; beer; Saccharomyces;		
XX	KM	fermentation; ds.		
XX	OS			
XX	Saccharomyces	cerevisiae.		
XX	Key	Location/Qualifiers		

DR	WP1; 1999-132448/11.
XX	
P1	New isolated cancer associated nucleic acids and polypeptides -
P7	isolated using sera from cancer patients, used to develop products
P7	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Claim 67; Page 742; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
S0	Sequence 1095 BP; 254 A; 269 C; 113 G; 432 T; 27 other;
Query Match	13.6%; Score 33; DB 20; Length 1095;
Best Local Similarity	54.5%; Pred. No. 0.62;
Matches	66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
OY	38 tctgtcgcgttcgcttcttcctccagatccctaagccagtgtgctaagaagcg 97
Db	7 tcccttcctcctctctctctctctctctctctctctctctctctctctacccaatcgc 66
OY	98 actaagactcttcgttcgcttgcttcctccgcgcagatcgtcgttgttgctcg 157
Db	67 actgctccctggaacctttccatacttcaccagaactgctgtgctgctgctct 126
OY	158 c 158
Db	127 c 127
RESULT 11	
N70435/c	
ID N70435 standard; cDNA; 1136 BP.	
XX N70435:	
AC	
DT 05-APR-1991 (first entry)	
DE Sequence encoding Insulin-like growth factor 1B (IGF-1B).	
KW Growth promoter; lactation enhancer; cell proliferation; ss.	
OS Homo sapiens.	
PN EP229750-A.	
PD 22-JUL-1987.	
PF 06-JAN-1987; 87EP-0870001.	
PR 20-NOV-1986; 86US-0929671.	
PR 07-JAN-1986; 86US-0816662.	
PA (UNIW) UNIV OF WASHINGTON.	
PI Krtvl GG, Rotweln PS;	
DR WP1; 1987-200203/29.	
XX New pre-pro-insulin-like growth factor-1 protein - obtd. by	

```

PI recombinant DNA procedures for use as growth promoters for
PI enhancing lactation, for stimulating cell proliferation etc.
XX
XX Example: Fig 5; 59pp; English.
XX
CC A 42 base oligonucleotide corresponding to the DNA sequence encoding
CC amino acids 10 to 23 of mature human IGF-I was synthesized (N70437).
CC The radiolabeled 42 mer was then employed to screen for IGF-I
CC containing DNA sequences in a human liver cDNA library. Insulin-
CC like growth factors-1A and -1B cDNAs were isolated from a human cDNA
CC library by using lambdagt11 (N70435, N70436). The human IGF-1
CC genomic gene was isolated and mapped. It encodes at least two
CC preproinsulin-like growth factor-1 proteins. An essentially pure
CC proproinsulin-like growth factor-1 protein comprising the sequence
CC of amino acids shown in Figure six is claimed (P70277).
XX
SQ Sequence 1136 BP; 412 A; 230 C; 268 G; 226 T; 0 other:

Query Match 13.6%; Score 32.8; DB 8; Length 1136;
Best Local Similarity 52.1%; Pred. No. 0.71;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 103 agctctgctgctgctgtctcttcctccgctcagatcgtctgctgctgctgata 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 727 ATCTCCCTCCCTCCCTGCTCTTCTCTTCCCTCGATCTGACACTGCTCTGCTCCCTCC 668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 163 accctaccctgctgctgctgctgctgctgctgctgctcctcctaagtcgatcga 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 TTCTCTTGTCCCTCCCTCGATGATGCTCTTGTGGCCACCTTTCCTTCTGTGAGACTTCTGTTC 608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 223 gtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 TTGTGTGATGATGGGGGCTG 588

RESULT 12
ID N50149
ID N50149 standard; cDNA; 3336 BP.
XX
XX N50149;
XX
XX 27-SEP-1991 (first entry)
XX
XX Sequence of the cDNA clone 23B6P.3 encoding a polypeptide
DE exhibiting mammalian immunoglobulin binding factor activity (IBF).
XX
XX Immunoglobulin E-mediated disease; therapy; B-cell differentiation;
KW Immunoglobulin E-binding factor; ss.
XX
XX Key Location/Qualifiers
FH CDS 94..1767
FT /*tag= a
XX
XX EPI55192-A.
XX
XX 18-SEP-1985.
XX
XX 15-MAR-1985; 85EP-0301834.
XX
XX 16-MAR-1984; 84US-0590430.
XX
XX (SCHE-) SCHERING BIOTECH CO.
PA (UYJO ) JOHNS HOPKINS UNIV.
PA (SCHE ) SCHERING-BIOTECH CO.
PA (DNAX-) DNAX RES INST MOLEC.
XX
XX Martens CL, Ishizaka K, Moore KM, Huff TF;
XX
XX WPI: 1985-231863/38.
XX
XX P-PSDB; P50120.
XX
XX New complementary DNA clones coding for poly(peptide)s - with

```

promoter binding DNA polymerase, which has a molecular weight

sequence 23532 B; 0510 A; 0141 C; 0514 G; 1013 T; 0 0line;

Query Match	13.18;	Score 31.8;	DB 19;	Length 29392;
-------------	--------	-------------	--------	---------------

OY	131	ctcaagatcgtctgcttgccttcgcttgacgaaccactcgtgctcgagttgcgctg	190
Dd	62	CTCAAGATCTGCTTGCCTTCCTCCTCGCTAGAACCCTACTGTCTCCGAATGTCGCTG	121
OY	191	cttgcgtctccctccccaagtcatgcattgltgtgtgagggggagcag	242
Dd	122	CTTGCTTTCTCCCTCCTCAAGTTCGATTCGATTGTGTGTGGGGGGCGCAG	173

RESULT 2

```

US-08-750-007-1
; Sequence 1, Application US/08750007
; Patent No. 5801016
; GENERAL INFORMATION:
; APPLICANT: MORIOKA, SHINJI
; TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT VECTOR
; TITLE OF INVENTION: CONTAINING THE SAME AND METHOD FOR EXPRESSING FOREIGN GENES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,007
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0760-221P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-750-007-1
    
```

Query Match 71.1%, Score 172; DB 1; Length 183;
 Best Local Similarity 100.0%; Pred. No. 8.5e-45;
 Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	71	ttaagccagcttgtcttagcctaagagcacctgatcttctgctgcgtcttcttcgcg	130
Dd	7	TAAAGCCAGctgtgctTAAGGCTAAGGCGACtAgGctTtTgtGctGctTtTCTTCG	66
OY	131	ctcagaatcgtctgcttgccttcgcttcgcttcgaacacctactcgtgctcgagttgcgctg	190
Dd	67	CTCAAGATCTGCTTGCCTTCCTCCTCGCTAGAACCCTACTGTCTCCGAATGTCGCTG	126
OY	191	cttgcgtctccctccccaagtcatgcattgltgtgtgagggggagcag	242
Dd	127	CTTGCTTTCTCCCTCCTCAAGTTCGATTCGATTGTGTGTGGGGGGCGCAG	178

RESULT 3

US-08-446-794A-5

```

: Sequence 5 Application US/08446794A
: Patent No. 5747327
:
: GENERAL INFORMATION:
: APPLICANT: Ueki, JUN
: APPLICANT: MORIOKA, SHINJI
: TITLE OF INVENTION: PHOSPHOLIPASE D GENE ORIGINATED FROM PLANT
: TITLE OF INVENTION: PLANT
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
: STREET: P.O. BOX 747
: CITY: FALLS CHURCH
: STATE: VA
: COUNTRY: USA
: ZIP: 22040-0747
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/446,794A
: FILING DATE:
:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: MURPHY JR, GERALD M
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 0760-0203P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2799 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1876..1983
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2524..2799
:
US-08-446-794A-5

```



```

1 TITLE OF INVENTION: GENES USING THE SAME
2 NUMBER OF SEQUENCES: 19
3 CORRESPONDENCE ADDRESS:
4 ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
5 STREET: PO BOX 747
6 CITY: FALLS CHURCH
7 STATE: VA
8 COUNTRY: USA
9 ZIP: 22040-0747
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.30
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/750,007
17 FILING DATE:
18 CLASSIFICATION: 435
19 ATTORNEY/AGENT INFORMATION:
20 NAME: MURPHY JR, GERALD M
21 REGISTRATION NUMBER: 28,977
22 REFERENCE/DOCKET NUMBER: 0760-221P
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (703) 205-8000
25 TELEFAX: (703) 205-8050
26 INFORMATION FOR SEQ ID NO: 4:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 2799 base pairs
29 TYPE: nucleic acid
30 STRANDEDNESS: single
31 TOPOLOGY: Linear
32 MOLECULE TYPE: DNA (genomic)
33 FEATURE:
34 NAME/KEY: CDS
35 LOCATION: 1876..1983
36 FEATURE:
37 NAME/KEY: CDS
38 LOCATION: 2524..2799
39 IS-08-750-007-4

```

RESULT 5
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA

```

COUNTY : USA
ZIP : 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER : IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE : PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300,6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE : (703)836-9300
TELEFAX : (703)683-4109
TELEX : 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pL-F15
US-08-232-463-14

Query Match 18.5%; Score 44.8; DB 1; Length 7218;
Best Local Similarity 6.0%; Pred.No.7e-05;
Matches 13; Conservative 129; Mismatches 76; Indels 0; Gaps 0;

OY 1 gtaagccagtgcttgctgaagcgacactagagctctgctgctgctcttc 60
Db 1045 GCACGCTGCAGGAGCTTGCGATTTTTTTTTTTTTTTTTTTTTTTTTT 1104

OY 61 gctagatcccaagccacgltgtcttagcctaagcgacactagagctctgctgtgc 120
Db 1105 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1164

OY 121 tctcttcgcgtcagatctgctgctgctgctgctgctaaccctactctgtgtcg 180
Db 1165 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1224

OY 181 agtgcgtctgctctgctctctctctcctcctaagtcgaltc 218
Db 1225 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1262

RESULT 6
US-08-931-999-4/c
Sequence 4, Application US/08931999
Patent No. 6043219
GENERAL INFORMATION:
APPLICANT: Iandolo, John J.
APPLICANT: Crupper, Scott S.
TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hovey, Williams, Timmons & Collins
STREET: 2405 Grand Boulevard, Suite 400
City: Kansas City
STATE: Missouri
COUNTRY: U.S.A.
ZIP: 64108
COMPUTER READABLE FORM:

```



```

;
;   NUMBER OF SEQUENCES: 12
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/515,604
;   FILING DATE: 27-APR-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 468,766
;   FILING DATE: 23-JAN-1990
;   APPLICATION NUMBER: 468,766
;   FILING DATE: 23-JAN-1990
;   APPLICATION NUMBER: 369,625
;   FILING DATE: 21-JUN-1989
;   APPLICATION NUMBER: 272,243
;   FILING DATE: 16-NOV-1988
;   APPLICATION NUMBER: 229,178
;   FILING DATE: 05-AUG-1988
;   APPLICATION NUMBER: 226,421
;   FILING DATE: 29-JUL-1988
;   APPLICATION NUMBER: 140,036
;   FILING DATE: 31-DEC-1987
;
; SEQ ID NO:1
;
;   LENGTH: 222
;
5274075-1
```

```

Query Match          11.7%; Score 28.2; DB 5; Length 222;
Best Local Similarity 53.1%; Pred. No. 2.7;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

QY 19 ggcgaagcgacactagagctctctgctgctctctccgcagatccaaagccca 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46 ggcgaagcgacactagagctctctgctgctctctccgcagatccaaagccca 105
QY 79 gtcgtcttaagcctaaagcctaaagctctctgctgctgctctctccgc 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 gtgagctacagcgccctcctacgctcctcatggtgcagcgctcctcctcgaac 158
```

```

RESULT 12
5274075-3
; APPLICANT: CHANG, TSE W.
; TITLE OF INVENTION: NEWLY IDENTIFIED HUMAN EPSILON
; IMMUNOGLOBULIN PEPTIDES AND RELATED PRODUCTS
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/515,604
; FILING DATE: 27-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 468,766
; FILING DATE: 23-JAN-1990
; APPLICATION NUMBER: 468,766
; FILING DATE: 23-JAN-1990
; APPLICATION NUMBER: 369,625
; FILING DATE: 21-JUN-1989
; APPLICATION NUMBER: 272,243
; FILING DATE: 16-NOV-1988
; APPLICATION NUMBER: 229,178
; FILING DATE: 05-AUG-1988
; APPLICATION NUMBER: 226,421
; FILING DATE: 29-JUL-1988
; APPLICATION NUMBER: 140,036
; FILING DATE: 31-DEC-1987
;
; SEQ ID NO:3
;
;   LENGTH: 378
;
5274075-3
```

```

Query Match          11.7%; Score 28.2; DB 5; Length 378;
Best Local Similarity 53.1%; Pred. No. 3.3;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

Db 202 ggcgaagcgccgctggaagctggaacgcgctctgcatcttcgcccgaactctccgtcctcagc 261
QY 79 gtcgtcttaagcctaaagcctaaagctctctgctgctgctctctccgc 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 gtgagctacagcgccctcctacgctcctcatggtgcagcggtcctcctcagc 314
```

```

RESULT 13
US-08-592-126-93/c
; Sequence 93, Application US/08592126
; Patent No. 5821091
;
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0960
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: G310con.seq
;
US-08-592-126-93
```

```

Query Match          11.3%; Score 27.4; DB 1; Length 1950;
Best Local Similarity 55.9%; Pred. No. 10;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```

QY 67 atccaaagccacgtgtgctcctaaagcctaaagcctctctgctcggtctctctc 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 997 accctgaccacagacttcaagaatagcagcactagcgcttgaactgtctctgc 938
QY 127 tccgctcagatctgctctgctctgctcgc 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 937 cACAGCGACATCGGGGATGCTCAGCTTGCT 905
```

```

RESULT 14
US-08-072-610-1/c
; Sequence 1, Application US/08072610
; Patent No. 5531133
;
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 21:02:13 ; Search time 1923.53 Seconds
(without alignments)
881.615 Million cell updates/sec

Title: US-09-600-602-2

Perfect score: 242
Sequence: 1 gtaagccagtgctgcttagg.....gtgtgtgtggtgggggcgcag 242

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743858 residues
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp14:*
 121: em_estp15:*
 122: em_estp16:*
 123: em_estp17:*
 124: em_estp18:*
 125: em_estp19:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est67:*
 136: gb_est68:*
 137: gb_est69:*
 138: gb_est70:*
 139: gb_est71:*
 140: gb_est72:*
 141: gb_est73:*
 142: gb_est74:*
 143: gb_est75:*
 144: gb_est76:*
 145: gb_est77:*
 146: gb_est78:*
 147: gb_est79:*
 148: gb_est80:*
 149: gb_est81:*
 150: gb_est82:*
 151: gb_est83:*
 152: gb_est84:*
 153: gb_est85:*
 154: gb_est86:*
 155: gb_est87:*
 156: gb_est88:*
 157: gb_est89:*
 158: gb_est90:*
 159: gb_est91:*
 160: gb_est92:*
 161: gb_est93:*
 162: gb_est94:*
 163: gb_est95:*
 164: gb_est96:*
 165: gb_est97:*
 166: gb_est98:*
 167: gb_est99:*
 168: gb_est100:*
 169: gb_est101:*
 170: gb_est102:*
 171: gb_est103:*
 172: gb_est104:*
 173: gb_est105:*
 174: gb_est106:*
 175: gb_est107:*
 176: gb_est108:*
 177: gb_est109:*
 178: gb_est110:*
 179: gb_est111:*
 180: gb_est112:*
 181: gb_est113:*
 182: gb_est114:*
 183: gb_est115:*
 184: gb_est116:*
 185: gb_est117:*
 186: gb_est118:*
 187: gb_est119:*
 188: gb_est120:*
 189: gb_est121:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	42.6	17.6	1201	190	CNS0167A	AL106384 Drosophila
2	37.2	15.4	668	94	AM694929	AM694929 NF081E10S
3	37.2	15.4	671	94	AM694702	AM694702 NF079B05S
4	36.8	15.2	733	139	BF036458	BF036458 601460209
5	36.2	15.0	1078	190	CNS0004P	AL075925 Drosophila
6	36	14.9	997	190	CNS005TE	AL060767 Drosophila
7	35.8	14.8	852	191	CNS02P31	AL194518 Tetradon
8	35.8	14.8	974	137	BE880847	BE880847 601493284
9	35.4	14.6	897	94	AM728159	AM728159 CA_pa001
10	35.2	14.5	737	174	AZ031415	AZ031415 RPT-23-2
11	35.2	14.5	746	107	BE415895	BE415895 MUG002.B0
12	35	14.5	404	143	N61793	N61793 T9ESTy2891
13	35	14.5	743	109	BE519618	BE519618 HV_CEB001
14	35	14.5	1235	137	RE881061	RE881061 601492122
15	34.8	14.4	490	22	A1563663	A1563663 yv01a11.x
16	34.8	14.4	937	190	CNS006ST	AL065880 Drosophila
17	34.4	14.2	632	97	AM974965	AM974965 EST387070
18	34.4	14.2	846	106	BE287958	BE287958 601094475
19	34.4	14.2	872	176	A2184405	A2184405 SP_1002_B
20	34.2	14.1	352	13	AA920986	AA920986 yv16h09.r
21	34.2	14.1	715	87	AM174924	AM174924 f130c07.y
22	34.2	14.1	990	193	CNS04YVF	AL134766 Tetradon
23	33.8	14.0	441	9	AA562562	AA562562 v172a07.r
24	33.8	14.0	588	178	AZ337956	AZ337956 1M0068B13
25	33.8	14.0	595	172	AQ928561	AQ928561 RPT-23-2
26	33.8	14.0	702	106	BE268416	BE268416 601124712
27	33.8	14.0	915	135	BE798838	BE798838 601584668
28	33.8	14.0	1101	190	CNS017GS	AL108022 Drosophila
29	33.6	13.9	462	182	AZ471215	AZ471215 1M0285N10
30	33.6	13.9	488	110	BE638959	BE638959 946017C08
31	33.6	13.9	520	177	A2258776	A2258776 RPT-23-1
32	33.6	13.9	544	2	AA071714	AA071714 mm71c09.r
33	33.6	13.9	586	26	A1881474	A1881474 606069E10
34	33.6	13.9	586	109	BE553696	BE553696 ur41a10.x
35	33.6	13.9	659	182	A463907	A463907 1M0273K03
36	33.4	13.8	443	16	A1121302	A1121302 uc30b03.r
37	33.4	13.8	584	27	A1981330	A1981330 pat.pk005
38	33.4	13.8	1203	190	CNS015YR	AL106077 Drosophila
39	33.4	13.8	1265	139	BR032829	BR032829 601455658
40	33.2	13.7	1101	190	CNS016U2	AL107204 Drosophila
41	33.2	13.7	1101	190	CNS0173D	AL107539 Drosophila
42	33.2	13.7	1101	190	CNS0181N	AL108773 Drosophila
43	32.8	13.6	473	10	AA703492	AA703492 zj11h02.s
44	32.8	13.6	481	89	AM385721	AM385721 PM3-LT003
45	32.8	13.6	618	28	AJ398711	AJ398711 AJ398711

ALIGNMENTS

RESULT 1
 LOCUS CNS0167A 1201 bp DNA GSS 26-JUL-1999
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
 BACN15C02 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESION AL106384
 VERSION AL106384.1 GI:5621653
 KEYWORDS GSS.

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
fruit fly.	Plasmodium falciparum	1 (bases 1 to 1201)	Genoscope.	Submitted		
Genoscope.				Direct Submission		
Submitted				(22-JUN-1999)	Genoscope	- Centre National de Sequencage
BP 191 91006 EVRY cedex	- FRANCE	(E-mail : seque@genoscope.cns.fr				
- Web : www.genoscope.cns.fr						
Determination of this BAC-end sequence was carried out as part of a						
collaboration with the European Drosophila Genome Project (EDGP) -						
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC						
library (Dros BAC) was made by Alain Billard at CEPH (Centre						
d'Etude du Polymorphisme Humain) with funding provided by a MRC						
project grant. The DNA was prepared from embryos by Alain Bucheton						
and Genevieve Payan. It has been constructed in the vector						
pBelobAC11.						
FEATURES	Source	Location/Qualifiers				
1..1201						
/organism="Drosophila melanogaster"						
/plasmid="pBelobAC11"						
/db_xref="taxon:7227"						
/clone_lib="DrosBAC"						
/clone="BACNISC02"						
/note="end : 77"						
BASE COUNT	238 a	261 c	148 g	261 t	293 others	
ORIGIN						
Query Match	17.6%	Score 42.6;	DB 190;	Length 1201;		
Best Local Similarity	28.0%;	Pred. No. 0.11;				
Matches 59;	Conservative 66;	Mismatches 86;	Indels 0;	Gaps 0;		
OY	30	ctaaagctcttgctgcggttgccttcttcgcctagatccaaagccagtgcttagg	89			
DB	965	KTGTGCTTTTCGCTKTTGTTCBCCCTSSCGTGSBCBCCSCCYGYSBGKCCGTG	1024			
OY	90	ctaaagcagctagagctcttctgcctgctgctcttccttcgcctagatctgcttgctg	149			
DB	1025	YGGGSGSYSCRTKCGCTGGGTGCTGCKKRYGTGCTSSSCSSCSGSKBKKKRYKKTKK	1084			
OY	150	ttgcttgctagagacctactctgtgctgcagagtgctgctgcttccttcctcaaa	209			
DB	1085	CBGSCGCCGCGSCCGCTCCYCBFTTCTCTYKTCYCTGTYGYSKTYCTTYGCTC	1144			
OY	210	gttcgactctgattgtgtgtgtgtggtggggcgc	240			
DB	1145	YGTGCKBCKSCGSKSTYBCTSYKCGTGYGC	1175			
RESULT 2						
LOCUS	AM694929/c	668 bp	mRNA	EST	15-JUN-2000	
DEFINITION	NE081E10STF1082	Developing stem	Medicago truncatula	CDNA clone		
ACCESSION	AF081E10ST	5', mRNA sequence.				
VERSION	AM694929					
KEYWORDS	AM694929.1	GI:7569691				
SOURCE	EST.					
ORGANISM	barrel medic.					
	Medicago truncatula					
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta					
	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;					
	Fabales; Fabaceae; Papilionoideae; Medicago.					
REFERENCE	1 (bases 1 to 668)					
AUTHORS	He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell					
	,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon					
	R.A.					
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation					
JOURNAL	Medicago truncatula stem library					
COMMENT	Unpublished (2000)					
	Contact: Dixon RA					

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel.: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert length: 668 Std Error: 0.00
Plate: 081 row: E column: 10
Seq primer: TCACACAGAAACAGACTATGAC.
Location/Qualifiers

FEATURES
source
1..668
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF081E10ST"
/clone_id="Developing stem"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/note="Vector: lambda zap; Contains a mixture of
internodal stem segments"

BASE COUNT
330 a 62 c 188 g 88 t

ORIGIN

Query Match 15.4%; Score 37.2; DB 94; Length 668;
Best Local Similarity 59.4%; Pred. No. 2.9;
Matches 63; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 117 ttgtctcttcgcgtagcatcgtcttgcttgctcgtagaccactctgtgc 176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 TTCCTGCACACGCCGTCCGTTCTGGTTCCTTTCATCTTCACTTATTCTTACTATTGT 373

OY 177 lgcagagtgcgctgcctgcctccctccaagltcgatcgtatt 222
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 TTCACTTGCTTCTTCATCTCCTTGCTACTCTTTCGCGTAATT 327

RESULT 3
AM694702/c 671 bp mRNA EST 15-JUN-2000
LOCUS
DEFINITION
NF079B05ST.F1044 Developing stem Medicago truncatula cDNA clone
NF079B05ST 5', mRNA sequence.

ACCESSION
AM694702
VERSION
AM694702.1 GI:7569464
KEYWORDS
EST.

SOURCE
barrel medic.
ORGANISM
Medicago truncatula
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 671)
He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
,C.J., Flores,H.R., Imman,J.T., Weller,J.W., May,G.D. and Dixon
,R.A.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula stem library
Unpublished (2000)

JOURNAL
COMMENT
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel.: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert length: 671 Std Error: 0.00
Plate: 079 row: B column: 05
Seq primer: TCACACAGAAACAGACTATGAC.
Location/Qualifiers
1..671
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF079B05ST"
/clone_id="Developing stem"
/tissue_type="stem"

FEATURES
source

	BASE COUNT	314 a	80 c	180 g	97 t
ORIGIN					
Query Match 15.4%: Score 37.2; DB 94; Length 671;					
Best Local Similarity 59.4%: Pred. No. 2.9; Indels 0; Gaps 0;					
Matches 63; Conservative 0; Mismatches 43;					
QY	117	tctgtctcttcgcgagatcgtctgtctgtctgtcgtaagaacctactctgtgc	176		
Db	272	TTCCTTCCGACATCGCTTCGGTTCTGTCTCCTTCTTCATTCTCTTATCTTACTATTOT	213		
QY	177	tgcgagtgctgcgtctgtctgtcttccctcccaagtcatctgatc	222		
Db	212	TTCAGTCTGTTCTTCATTCTCTCTGCTTACTCTTTTCGCCGTAAATT	167		
RESULT 4					
BFO36458/c		733 bp	mRNA	EST	10-OCT-2000
LOCUS	BFO36458	601460209F1 NIH_MGC_66	Homo sapiens	cDNA clone IMAGE:3863690 5'	
DEFINITION		mRNA sequence.			
ACCSSION	BFO36458				
VERSION	BFO36458.1	GI:10744499			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
TITLE	1 (bases 1 to 733)				
JOURNAL	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Tel: (301) 496-1550				
	Email: Robert_Strausberg@nih.gov				
	Tissue Procurement: DCDP/DTP				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: LCM591 row: a column: 03				
	High quality sequence stop: 685.				
FEATURES					
Source					
	1..733				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:3863690"				
	/clone_lib="NIH_MGC_66"				
	/tissue_type="adenocarcinoma"				
	/lab_host="DH10B (phage-resistant)"				
	/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."				
BASE COUNT	175 a	195 c	208 g	155 t	
ORIGIN					
Query Match 15.2%: Score 36.8; DB 139; Length 733;					
Best Local Similarity 61.5%: Pred. No. 3.7;					
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;					
QY	101	agaacattcttcgcgtcgtctcttcctccggccaatctgtctgtctgtcgcga	160		
Db	729	AGAGCAGGACGCGCTTTTCCTCTCCTCCGCTGCGCACTTGCGGCTTCTGTCTTCCCT	670		
QY	161	gaaccctactctgtgtcgtcgaagtgcgtctgtcgt	196		

D6	669	GACCCCTCCTCTTCCGCGGTGTCGTCCTTTTAGTG	634
<hr/>			
RESULT	5		
CNS00J4P			GSS
LOCUS			03-JUN-1999
DEFINITION			
Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR37E1S of RpCl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1078)			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RpCl-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.			
FEATURES			
SOURCE			
1..1078 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1id="RpCl-98" /clone="BACR37E1S" /note="end : TET3"			
BASE COUNT			
ORIGIN			
187 a 210 c 251 g 272 t 158 others			
<hr/>			
Query Match			
Best Local Similarity	46.5%:	Pred. No. 5.8:	
Matches	60;	Conservative	16; Mismatches 53; Indels 0; Gaps 0;
OY	112	cctgcgtccttcctccgcatcgatcgttgcgttcgttcgttcgttcgaagaccctactc	171
		
D6	525	CTGGCVTTCTTKGTTBGTTWCCCCGMMTTTTCTCTGKGGMWBCHBTATCYCTAMTT	584
OY	172	tgtctcgtcgaattgcgttcgttcgttccttcctcctaagtcatcgatgttgtagtg	231
		
D6	585	TWARKKGYTTTATGAVNCVCTTCGTGCTKCTCTGCTGATTGTCTCGAKSGASTCKXTG	644
OY	232	ggaggagggcgc	240
		. . .	
D6	645	GGGGGW\$AC	653
<hr/>			
RESULT	6		
CNS005TE			
LOCUS			
DEFINITION			
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR12K22 of RpCl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			

Db	120	GCCTCTGTCCTTGGTGATAGCGCGTCGCCGACACC	86
RESULT	13		
LOCUS	BE519618		
DEFINITION	BE519618	743 bp mRNA	EST 08-AUG-2000
VERSION	HV_CEBD0016H02f	Hordeum vulgare seedling green leaf EST library	
SOURCE	HVCEDN0005	(Erysiphe infected & control)	Hordeum vulgare cDNA clone
ORGANISM	BE519618		
REFERENCE	BE519618		
AUTHORS	BE519618.1	GI:9743722	
TITLE	Development of a genetically and physically anchored EST resource for barley genomics		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Wing RA Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Tel.: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: AATTAACTCCATCAAGAAGG High quality sequence start: 148 High quality sequence stop: 584. Location/Qualifiers 1..743 /organism="Hordeum vulgare" /cultivar="CI16151 (ML6)" /db_xref="taxon:4513" /clone="HV_CEBD0016H02f" /clone_1lb="Hordeum vulgare seedling green leaf EST library HVCEDN0005 (Erysiphe infected & control)" /tissue_type="seedling green leaf" /lab_host="SOLR" /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	24 a	256 c	88 g 363 t 12 others
ORIGIN			
Query Match	14.5%	Score 35;	DB 109; Length 743;
Best Local Similarity	55.3%	Pred. NO. 12;	
Matches	68; Conservative	0; Mismatches	55; Indels 0; Gaps 0;
OY	105	ctctgtcgtcgttcctctctctccgcacagatcgctcgtctgtcgttcgtcagaac	164
Db	392	CTTGTCGTTCGGCGCTTTCTTCTCTGTTACCTCCGTTGTGTTCTGTCTCTTAATGCTC	451
OY	165	cctaactcgtgctgcgaagtgctgcgtcgtcgtcctcctcccaagtcatgatcgt	224
Db	452	CCTGCTTGTCTCGTCGTCGTCCTTCTGTCGTCCTTGTGTCGTCGTCGTTTCCTTT	511
OY	225	gtc 227	
Db	512	TTG 514	
RESULT	14		
LOCUS	BE881061/c		
DEFINITION	BE881061	1235 bp mRNA	EST 27-SEP-2000
VERSION	601492122F1	NIH_MGC_69 Homo sapiens CDNA clone IMAGE:3894292	5',
ACCESION	BE881061		
VERSION	BE881061.1	GI:10329837	

	KEYWORDS	EST.	human.
SOURCE	ORGANISM	Homo sapiens	
REFERENCE	AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE	JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
COMMENT		1 (bases 1 to 1235)	
		NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/ .	
		Unpublished Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)	
		Contact: Robert Strausberg, Ph.D.	
		Tel.: (301) 496-1550	
		Email: Robert.Strausberg@nih.gov	
		Tissue Procurement: DCRD/DMP/Gazar	
		cDNA Library Preparation: Life Technologies, Inc.	
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)	
		DNA Sequencing by: Incyte Genomics, Inc.	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LNL at:	
		http://image.llnl.gov	
		plate: LNC670 row: 1 column: 05	
FEATURES	source	High quality sequence stop: 217.	
		Location/Qualifiers	
		1..1235	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:3894292"	
		/clone_1lb="NH_MGC_69"	
		/tissue_type="large cell carcinoma, undifferentiated"	
		/lab_host="DH10B (phage-resistant)"	
		/note="Organ: Lung; Vector: pCMV-SPORT6; Site:1; NotI;	
		Site:2; SalI; Cloned unidirectionally. primer: Oligo dT."	
		Average insert size 1.1 kb. Library constructed by Life	
		Technologies."	
BASE COUNT	723 a	61 c	373 g 78 t
ORIGIN			
Query Match	14.5%	Score 35;	DB 137; Length 1235;
Best Local Similarity	50.3%;	Pred. NO. 12;	
Matches	86;	Conservative	0; Mismatches 85; Indels 0; Caps 0
OY	36	cttcctgctcgttgtcttcctccgcagatcctaaccagtgcttagctaagc	95
Db	835	CTGCCCTCTCGTGTCCTTCTCCTCCGCTCCGCTCTCTCTGTCTTTCCCTTCTCTCTTC	776
OY	96	gcacatagaagctctctcgcgcgtcgtctcttcctccgcactagaatcgtctgctgt	155
Db	775	TGTTTCTGCTCTTCTTTTCTCTCCGCTTCGGCTCTTCGCAACGCCCTTTTCTCTTTCCCT	716
OY	156	cagtagaacctactctgtgctcgagagtgcgtcgtctcgtctctcctct	206
Db	715	TCCTAGATTCCTCTTTCCTCTTAATTTCTCTTCTTTGTGTGTCGTCCT	665
RESULT 15			
LOCUS	AI563663/c		
DEFINITION	AI563663	490 bp mRNA	EST 26-MAR-1999
ACCESSION	AY01411.x1	Stratogene mouse macrophage (#937306)	Mus musculus
VERSION	clone IMAGE:1294172.3	similar to gp:M10062 Mouse IgE-binding	CDNA
KEYWORDS	factor mRNA, complete cds (MOUSE);,	mRNA sequence.	
SOURCE	AI563663.1 GI:4522120		
ORGANISM	EST.	house mouse.	
		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
		1 (bases 1 to 490)	
		Marra,M., Hillier,L., Kucaba,T., Martin,J.J., Beck,C., Wyllie,T.,	
		Underwood,K., Steptoe,M., Theising,B., Hallen,M., Bowers,Y., Person	
		,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Riltter	
		,E., Kohn,S., Shu,T., Jackson,Y., Cardenas,M., McCann,R.,	
		Waterson,R. and Wilson,R.	

TITLE The Nashu-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/Washu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:675220

This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 390.

FEATURES

source

Location/Qualifiers
1..490
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1294172"
/clone_lib="Stratagene mouse macrophage (#937306)"
/tissue_type="macrophage"
/dev_stage="WEHI-3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: blood; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
0.190 dt. WEHI-3 cell line. Average insert size: 1.5 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 136 a 110 c 125 g 119 t
ORIGIN

Query Match 14.4%; Score 34.8; DB 22; Length 490;
Best Local Similarity 60.6%; Pred. No. 12;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Oy 14 gcttagcctaagcgactagagctctgtcgtcgtctctctctcgcctcagatcctaa 73
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 GCTCTTGCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTGCG 67
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 74 gccacgtgtgtcttaggcctaagcgactagagctt 107
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 ACCCTGCTCCTGAGATGTAAGAAATAAGCTT 33

Search completed: February 12, 2001, 21:02:16
Job time: 3972 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: February 12, 2001, 21:27:36 ; Search time 1275.75 Seconds

(without alignments)
694.000 Million cell updates/sec

Title: US-09-600-602-3

Sequence: 1 gtaagcccaagtgtgcttaggttgtgtgtggggcgcgcaag 173

Scoring table: IDENTITY_NUC

Searched: 111813 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Pre-processing:	Minimum Match	08
Post-processing:	Minimum Match	08

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *

```
2: gbd_da: *
3: gbd_om: *
```

4: gbb_ov:*
5: gbb_ph:*

*:gb_p11:6:

8: gcb_pr1: *

```
9:  gb_pr2:
10:  qb_pr3:
```

```
11: gb_ro:*
12: gb_sv:*
```

13: qb_un: *

15: em_hum1

```
16: em_iniz
17: em_in: *
```

```
18: em_om: *
19: em_or: *
```

20: em_ov: *

22: em_ph: *

```
23: em_pl: *
24: em_ro: *
```

25: em_sts:
26: em_ev: #

27: em_un: *

29: gcb_htg1

```
30:  gb_nlgz
31:  qb_inl:
```

32: gb_in2:
33: am_ba1:

34: em_ba2:

36: em_hum4

37: qb-pr4:
38: qb-htq3

39: gb_hlg4
40: gb_hlg5

41: gcb_hcgc6

43: em_htg1

44:	em.itrg3:*
45:	em.itrg3:*
46:	em.hung3:*
47:	gd.pl13:*
48:	gd.pr13:*
49:	gd.itrg9:*
50:	gd.itrg9:*
51:	gd.itrg10:*
52:	gd.itrg11:*
53:	gd.itrg12:*
54:	gd.itrg13:*
55:	gd.itrg14:*
56:	gd.in3:*
57:	gd.itrg15:*
58:	gd.itrg16:*
59:	gd.itrg17:*
60:	em.itrg4:*
61:	em.itrg5:*
62:	em.itrg9:*
63:	em.itrg7:*
64:	em.hung7:*
65:	gd.itrg18:*
66:	gd.itrg19:*
67:	gd.itrg20:*
68:	gd.itrg21:*
69:	gd.itrg22:*
70:	gd.itrg23:*
71:	gd.v11:*
72:	gd.v12:*
73:	gd.daz3:*
74:	em.itrg8:*
75:	em.itrg3:*
76:	em.hung10:*
77:	gd.pr7:*
78:	gd.pr6:*
79:	gd.sts1:*
80:	gd.sts2:*
81:	gd.pat1:*
82:	gd.daz7:*
83:	em.itrg0:*
84:	gd.itrg4:*
85:	gd.pr8:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	173	100.0	173	81	AR037064	AR037064 Sequence
2	173	100.0	183	81	AR037061	AR037061 Sequence
3	173	100.0	2799	81	AR005013	AR005013 Sequence
4	173	100.0	2799	81	AR037063	AR037063 Sequence
5	173	100.0	5871	6	AB001920	AB001920 OriZya sat
6	36.4	21.0	146903	57	AC068648	AC068648 Homo sapi
7	35.6	20.6	196866	38	AC010742	AC010742 Homo sapi
8	35.4	20.5	93540	8	AB042235	AB042235 Homo sapi
9	35.4	20.5	126334	66	AL133403	AL133403 Homo sapi
10	35.4	20.5	182821	8	AB041340	AB041340 Homo sapi
11	35.4	20.5	181302	37	AL355532	AL355532 Human DNMT
12	35.4	20.5	201251	52	AC025964	AC025964 Mus musc
13	35.4	20.5	204865	53	AC026682	AC026682 Mus musc
14	35.4	20.5	300000	48	AF002530	AF002530 Homo sapi
15	35	20.2	201833	84	CMS07407	AF002530 Homo sapi
16	34.8	20.1	173661	59	AC057107	AL135102 Homo sapi
17	34.6	20.0	7218	81	I66494	AC074107 Mus musc
18	34.6	20.0	111507	30	AC010001	I66494 Sequence 14
19	34.6	20.0	167253	58	AC065949	AC010001 Mus musc
20	34.6	20.0	207757	51	AC023934	AC065949 Homo sapi
21	34.4	19.9	6167	7	CCPc1c1GE	AC023934 Homo sapi
						X89390 C.concinnia

22	34.4	19.9	154959	52	AC025125	AC025125 Homo sapi
23	34.2	11.8	95280	53	AC026770	AC026770 Homo sapi
C 24	34.2	19.8	166606	42	AC020685	AC020685 Homo sapi
25	34	19.7	116534	58	AC069368	AC069368 Homo sapi
26	33.8	19.5	326965	50	AC022126	AC022126 Homo sapi
C 27	33.6	19.4	534	80	HUMUT1934	L18059 Human STS U
C 28	33.6	19.4	82188	6	AB010076	AB010076 Arabidops
C 29	33.4	19.3	377	80	HUMUT846	L18446 Human STS U
C 30	33.4	19.3	39229	8	AC004755	AC004755 Homo sapi
C 31	33.4	19.3	163297	57	AC068847	AC068847 Homo sapi
C 32	33.2	19.2	1460	73	TRPBP	M17716 T.pallidum
C 33	33.2	19.2	14470	1	AE001269	AE001269 T.repennu
C 34	33.2	19.2	68418	65	AC079068	AC079068 Homo sapi
C 35	33.2	19.2	167994	39	AC012084	AC012084 Homo sapi
C 36	33.2	19.2	168953	38	AC011691	AC011691 Homo sapi
C 37	33.2	19.2	174487	44	AC012011	AC012011 Homo sapi
C 38	33.2	19.2	198211	59	AC073756	AC073756 Mus muscu
C 39	33.2	19.2	204504	41	AC016820	AC016820 Homo sapi
C 40	33.2	19.2	209920	59	AC035682	AC035682 Mus muscu
C 41	33.2	19.2	223193	51	AC024462	AC024462 Homo sapi
C 42	33	19.1	375782	31	AC009602	AC009602 Leishman
C 43	33	19.1	230760	59	AC073714	AC073714 Mus muscu
C 44	32.8	19.0	185	81	E01350	E01350 Synthetic T
C 45	32.8	18.5	81	108008	Sequence 1	I08008 Sequence 1

ALIGNMENTS

RESULT	1			
AR037064				
LOCUS	AR037064	173 bp	DNA	
DEFINITION	Sequence 7 from patent US 5801016.			
			PAT	29-SEP-1999

ACCESSION	AR037064	GI:5954920	
VERSION	AR037064.1		
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 173)		
TITLE	Morioka,S. and Ueki,J.		
JOURNAL	DNA fragment, recombinant vector containing the same		
FEATURES	expressing foreign genes using the same		
source	Patent: US 5801016-A 7 01-SEP-1998;		
	1..173		
	/organism="unknown"		
BASE COUNT	21 a 48 c 47 g 57 t		
ORIGIN			

[illegible]

RESULT	2				
AR037061					
LOCUS		183 bp	DNA		
DEFINITION		Sequence 1 from patent US 5801016.			
ACCESSION	AR037061			PAT	29-SEP-1999

VERSION	AR037061.1	GI:5954917	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 183)		
TITLE	Morioka,S. and Ueki,J.		
JOURNAL	DNA fragment, recombinant vector containing the same and method for		
FEATURES	expressing foreign genes using the same		
SOURCE	Patent: US 5801016-A 1 01-SEP-1998;		
	Location/Qualifiers		
	1..183		
	/organism="unknown"		
BASE COUNT	23 a	51 c	51 g
ORIGIN			58 t

	Query Match	Best Local Similarity	100.0%	Score 173	DB 81	Length 183
			100.0%	Pred. No. 1.2e-42		
	Matches 173	Conservative	0	Mismatches	0	Indels 0
				Gaps		
Qy	1	gtaagcccaagctgtccttagagcgaacagacagagctcttcgtcgcgtctctcttc	60			
Dd	6	GTAGCCCAAGTGTGCTTAGGCTAGGCCACTAGAGCTTCTTGTCTTCTTCTCC	65			
Qy	61	gtcaagalcctgtccttgccttgccttcgcgtlaagaccctactcctgtctgcaagtgatc	120			
Dd	66	GCTCAGATCTGCTTGCCTTGTGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT	125			
Qy	121	gtctcgtctctctctctcctcaagcttcgaactcgtatctgtctgtgagggggcgaag	173			
Dd	126	GCCTTCGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC	178			

[illegible]

	Query Match	Similarity	100.0%	Score 173:	DB 81:	Length 2799:
	Best Local	Similarity	100.0%	Pred. No. 1,3e-42:		
	Matches 173:	Conservative	0:	Mismatches 42:	Indels	Gaps
Qy	1	gtaaacccagtgctgcttagagcctaagcgcacagagctctcttcgcgtgctctctcc	60			
Db	1666	GTAAccCAGtGCTTgAGGCTAAAGCAGACTTCTTgCTTgCTTCTTCTTC	1725			
Qy	61	gtccagacatctgcttgctgctgctctgcagaaacctactctgctgctgagatgcgt	120			
Db	1726	GCTCAGATCTGCTTCTTGTGCTTGCCTGAACCTTACTCTGTGTCGAGTGTGCT	1785			
Qy	121	gcttcgctcttcctccctcaagtcagatctgattgtgtgtgtggtggggcgacg	173			
Db	1786	GCTTCGCTTCTTCCTTCCTCAAGTTCAGTCTGTGATGTGTGTGGGGGGCGCAG	1838			

RESULT 4


```
REFERENCE
AUTHORS
1 (bases 1 to 146903)
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodot,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunc,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Ferrante,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogue,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondzewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichter,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,B., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R.,
Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Webb,M.,
Wellington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Wolley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
TITLE
JOURNAL
Direct Submission
Unpublished
2 (bases 1 to 146903)
Mocley,K.C.
Direct Submission
Submitted (06-May-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 6, 2000 this sequence version replaced gi:9665001.
COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBBG
Center clone name: RP11-331C18
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 5% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 139132 bases at least Q40
Consensus quality: 143411 bases at least Q30
Estimated insert size: 145399; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
41225: contig of 41225 bp in length
41226: gap of unknown length
41325: gap of unknown length
71989: contig of 30664 bp in length
71990: gap of unknown length
72090: contig of 28171 bp in length
100260: contig of 28171 bp in length
100261: gap of unknown length
100361: contig of 23993 bp in length
124353: gap of unknown length
124354: gap of unknown length
124454: contig of 12849 bp in length
137402: gap of unknown length
137403: contig of 9501 bp in length.
Location/Qualifiers
1.146903
/organism="Homo sapiens"

/dB_xref="taxon:9606"
/chromosome="12"
/clone="RP11-331C18"
BASE COUNT 42841 a 28346 c 28517 g 46697 t 502 others
ORIGIN
Query Match 21.0%; Score 36.4; DB 57; Length 146903;
Best Local Similarity 59.8%; Pred. No. 1.1;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
Oy 37 ttctgtcgcgtctgtcttcttcgcgtcagatcgtctgtctgtctgtcagaac 96
Db 81368 TCTCTGCGCGTTCCTTCCTGCTCCGCTCCCTCCCTCTTCCTTCCTTCGCG 81447
Oy 97 ctactcgtctgcagatgcgtctgtctgtctgtctgtctccccc 138
Db 81448 TTCTTCTTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 81489

RESULT 7
AC010742/c
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-524015, WORKING DRAFT
ACCESSION
AC010742
VERSION
AC010742.3 GI:7631043
KEYWORDS
HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 196866)
Waterston,R.H.
TITLE
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 196866)
Waterston,R.H.
Direct Submission
Submitted (21-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 21, 2000 this sequence version replaced gi:6648301.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information
Center project name: H.NH0524015
----- Summary Statistics
Sequencing vector: pLasmid; %
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 194507 bases at least Q40
Consensus quality: 195304 bases at least Q30
Consensus quality: 195774 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 196466; sum-of-contigs
Quality coverage: 6.30 in Q20 bases; agarose-fp
Quality coverage: 6.74 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3370: contig of 3370 bp in length
```


JOURNAL Submitted (07-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hunkquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:9926594.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>
Rp11-487F5 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://Dacpac.med.buffalo.edu/VECTOR.pbacc3.6>

FEATURES

Source

1.181302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Rp11-487F5"
/clone.lib="RPCI-11.2"
165..209
repeat_region repeat: matches 6139..6185 of consensus"
210..575
repeat_region repeat: matches 1..364 of consensus"
576..1153
repeat_region repeat: matches 5595..6139 of consensus"
1167..1267
repeat_region repeat: matches 6075..6171 of consensus"
1268..1528
repeat_region repeat: matches 35..295 of consensus"
1529..2819
repeat_region repeat: matches 4748..6075 of consensus"
2940..4010
repeat_region repeat: matches 3502..4589 of consensus"
4033..4216
repeat_region repeat: matches 3..170 of consensus"
4238..4259
repeat_region repeat: matches 11 copies 2 mer to 100% conserved"
4264..4819
repeat_region repeat: matches 2857..3448 of consensus"
/note="L1M8 repeat: matches 4517..4588 of consensus"
4951..5020
repeat_region repeat: matches 5792..6150 of consensus"
5032..5389
repeat_region repeat: matches 5792..6150 of consensus"
/note="L1P83 repeat: matches 5792..6150 of consensus"
Complement(5266..5803)
/note="match: GSS: Em:A0286760"
Complement(5274..5515)
/note="match: GSS: Em:A0745266"
Complement(5285..5479)
/note="match: GSS: Em:A0884205"
Complement(5421..5672)
/note="match: GSS: Em:A0432202"

misc_feature complement(5498..5803)
repeat_region /note="match: GSS: Em:A0269249"
5504..7766
/note="L1M1 repeat: matches 3572..5793 of consensus"
5570..5802
misc_feature /note="match: GSS: Em:A0134658"
7855..8378
repeat_region repeat: matches 3014..3565 of consensus"
8393..9420
repeat_region repeat: matches 731..1693 of consensus"
9454..9713
repeat_region repeat: matches 144..409 of consensus"
10581..10997
repeat_region repeat: matches 144..409 of consensus"
11012..11744
repeat_region repeat: matches 669..242 of consensus"
/note="L1M4 repeat: matches 2111..2829 of consensus"
11794..11934
repeat_region repeat: matches 1888..2044 of consensus"
11945..12128
repeat_region repeat: matches 128..311 of consensus"
12130..12167
repeat_region repeat: matches 19 copies 2 mer to 78% conserved"
12179..12310
repeat_region repeat: matches 1..132 of consensus"
12456..12657
repeat_region repeat: matches 318..518 of consensus"
12651..12780
repeat_region repeat: matches 177..47 of consensus"
12969..13040
repeat_region repeat: matches 269..196 of consensus"
13050..13848
repeat_region repeat: matches 1..790 of consensus"
16774..17008
repeat_region repeat: matches 209..464 of consensus"
17242..17504
repeat_region repeat: matches 92..360 of consensus"
17520..18008
repeat_region repeat: matches 3145..3672 of consensus"
18846..19124
misc_feature /note="match: GSS: Em:A0371909"
18930..19345
repeat_region repeat: matches 74..486 of consensus"
19412..19494
repeat_region repeat: matches 1..79 of consensus"
20224..21691
repeat_region repeat: matches 4667..6157 of consensus"
21644..22028
misc_feature /note="L1P16 repeat: matches 4667..6157 of consensus"
21818..21957
repeat_region repeat: matches 40..178 of consensus"
22909..23035
repeat_region repeat: matches 2616..2745 of consensus"
23100..23484
repeat_region repeat: matches 1..397 of consensus"
23804..24316
repeat_region repeat: matches 1242..1814 of consensus"
24240..24717
misc_feature /note="match: GSS: Em:A0344549"
24264..24713
misc_feature /note="match: GSS: Em:A0267518"
24285..24803
misc_feature /note="match: GSS: Em:A05813"
24423..24518
repeat_region repeat: matches 48 copies 2 mer to 74% conserved"
25084..25371
repeat_region repeat: matches 3..295 of consensus"
25938..26382
repeat_region repeat: matches 9..458 of consensus"
27817..27981
repeat_region repeat: matches 192..362 of consensus"
28454..28515
repeat_region repeat: matches 2648..2709 of consensus"
Complement(29112..29766)
misc_feature


```

AUTHORS
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Deterlich, D., Thomas, S., Okunou, G., Carlack, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Bunay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kosay, C., Liu, J., Liu, W., Louised, H., Lozado, R. J., Martin, R.,
Massey, E., McLeod, M. P., Mel, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Relter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.

TITLE
Direct Submission

JOURNAL
Unpublished

REFERENCE
2 (bases 1 to 204806)

AUTHORS
Worley, K. C.

TITLE
Direct Submission

JOURNAL
Submitted (23-MAH-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 5, 2000 this sequence version replaced gi:9930744.

COMMENT
----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: MANDI
Center clone name: RP23-201K2
----- Summary Statistics
Sequencing vector: M13, L08821
Chemistry: Dye-terminator Big Dye: 26% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 204424 bases at least Q40
Consensus quality: 210404 bases at least Q30
Consensus quality: 212657 bases at least Q20
Estimated insert size: 203906; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-ef estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 86612: contig of 86612 bp in length
* 86613 86712: gap of unknown length
* 86713 157534: contig of 70822 bp in length
* 157535 157634: gap of unknown length
* 157635 176497: contig of 18863 bp in length
* 176498 176597: gap of unknown length
* 176598 194222: contig of 17625 bp in length
* 194223 203328: gap of unknown length
* 203329 203498: contig of 9076 bp in length
* 203499 204806: gap of unknown length
* 204807 204806: contig of 1308 bp in length.

FEATURES
Source
1. 204806
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-201K2"

BASE COUNT
55914 a 47687 c 46801 g 53889 t 515 others

ORIGIN

```

[illegible]

THIS PAGE BLANK (USPTO)

ID		086785 standard; DNA; 2799 BP.
XX		
AC	O86785;	
XX		
DT	04-MAR-1996 (first entry)	
XX		
DE	DNA encoding Phospholipase D.	
XX		
KW	phospholipase D; measurement; reagent; phospholipid level; ss.	
XX		
OS	Zea mays.	
XX		
FH	Key	Location/Qualifiers
FT	exon	1876..1983
FT		/tag= a
FT	Intron	1984..2523
FT		/tag= b
FT	exon	2524..2799
FT		/tag= c
FT		/note= "partial exon 2"
XX		
PN	WO9509234-A1.	
XX		
PD	06-APR-1995.	
XX		
PF	30-SEP-1994; 94MO-JP01627.	
XX		
PR	30-SEP-1993; 93JP-0267884.	
XX		
PA	(NISB) JAPAN TOBACCO INC.	
XX		
P1	Morioka S, ueki J;	
XX		
DR	WPI; 1995-147433/19.	
DR	P-PSDB; R72799.	
XX		
PT	Cloned DNA coding plant derived phospholipase D - controls	
XX	expression of plant derived PLD gene	
XX		
PS	Claim 14; Page 33-35; 41pp; Japanese.	
CC	The DNA contains sequences necessary for the expression of a plant	
CC	derived phospholipase D (PLD). The PLD is useful for measuring	
CC	phospholipid levels and for producing derivs. by e.g. base exchange	
CC	reactions.	
XX		
SO	Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;	
<hr/>		
	Query Match	100.0%; Score 173; DB 16; Length 2799;
	Best Local Similarity	100.0%; Pred. No. 1.9e-44;
	Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 gtaagccagatgtcttgactgagcctaagcgacataagatcttgcctgtcttcctcc 60	
Db	1666 gtaagcccagatgtgctttagcctaagcgactagaactctctgcctgtcttcctcc 1725	
Oy	61 gctcagactctcttgcctgtctgcctgcctgaaccctactctgtcgtcgagttgct 120	
Db	1726 gctcagactctgtcgtctgtctgtcctgcctgaaccctactctgtcgtcgagttgct 1785	
Oy	121 gcttcgtcttccttcctcaagtctgatattgtgtgtggggggcgcag 173	
Db	1786 gcttcgtcttccttcctcaagtctgatattgtgtgtgtggggggcgcag 1838	
<hr/>		
RESULT	5	
T42854		
ID	T42854 standard; cDNA to mRNA; 2799 BP.	
XX		
XC	T42854;	
XX		
DT	16-JUN-1997 (first entry)	

XX		Phospholipase D gene sequence fragment.
DE		
XX	KW	Phospholipase D; rice; promoter; ss.
XX	OS	Oryza sativa.
PH	Key	Location/Qualifiers
FT	exon	1876..1983
FT		/tag= a
FT		/number= 1
FT		/note= "encodes residues 1 to 36 of W06134"
FT	intron	1984..2523
FT		/tag= b
FT		/number= 1
FT	exon	2524..2799
FT		/tag= c
FT		/number= 2
FT		/note= "encodes residues 37 to 128 of W06134"
XX		
PN		W09630510-A1.
XX		
PD		03-OCT-1996.
XX		
PE		28-MAR-1996; 96MO-JP00812.
PR		29-MAR-1995; 95JP-0096126.
XX		
PA		(NISB) JAPAN TOBACCO INC.
PI		Morioka S, Ueki J;
XX		
DR		WPI: 1996-455357/45.
DR		P-PSDB: W06134.
XX		
PT		Promoter DNA sequence derived from rice - used to increase expression of foreign genes in transformed hosts
PS		Disclosure; Page 20-22; 29pp; Japanese.
CC		This sequence represents a fragment of the coding sequence of the rice phospholipase D gene (PLD). The promoter for the PLD gene was isolated using the primers shown in T42857 and T42586. The promoters (see T42851 and T42852) are efficient promoters for greatly increasing the expression of foreign genes in transformant rice and other plants.
CC		
CC		
CC		
CC		
CC		
SO		Sequence 2799 BP; 692 A; 709 C; 609 G; 789 T; 0 other;
Query Match	100.0%; Score 173; DB 17; Length 2799;	
Best Local Similarity	100.0%; Pred. No. 1.9e-44;	
Matches 173; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Dy	1 gtaaacccagtgctgactgaagcgcacctaagacctgtgctgcgttctcttc 60	
Dd	1666 gtaaacgccagtgtgcttagcgtaaagcgcaactagaagcttctgctgccttctctcc 1725	
Oy	61 gtccgagatcgttgcttgcttgctgcttgctgtagaaacctactctgtgctgcagattgctgt 120	
Dd	1726 gtccgagatcgttgcttgcttgctgcttgctgtagaaacctactctgtgctgcagattgctgt 1785	
Oy	121 gtctgcttcttcctcctcaagttcgaatcgaatctgtgtgtgtggggggggcag 173	
Dd	1786 gtctgcttcttcctcctcaagttcgaatcgaatctgtgtgtgtgtggggggggcag 1838	
RESULT	6	
XI3856		
ID	XI3856 standard; DNA; 796 BP.	
XX		
AC	XI3856;	
XX		
DT	19-MAR-1999 (first entry)	

[illegible]

XX	Polynucleotide sequence from the genome of <i>Treponema pallidum</i> .
DE	
XX	<i>Treponema pallidum</i> infection; syphilis; Borrelia infection; animal;
KW	enzyme production; ds.
XX	
OS	<i>Treponema pallidum</i> .
XX	
PM	W09859034-A2.
XX	
PD	30-DEC-1998.
XX	
PF	23-JUN-1998; 98WO-US13041.
XX	
PR	24-JUN-1997; 97US-0050667.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Fraser CM;
XX	
DR	WPI: 1999-081273/07.
XX	
PT	New isolated <i>Treponema pallidum</i> nucleic acids - used to develop
PT	products for the detection, diagnosis, characterisation, prevention
PT	and therapy of <i>T. pallidum</i> infections, particularly syphilis
XX	
PS	Claim 1; Page 608-619; 1150pp; English.
XX	
CC	X20500-21243 represent polynucleotide sequences from the genome of
CC	<i>Treponema pallidum</i> . The sequences can be used for detection,
CC	diagnosis, characterisation, prevention and therapy for <i>T. pallidum</i>
CC	infections, particularly syphilis. They can also be used for detecting
CC	diseases related to Borrelia infections in animals, and for the
CC	production of biosynthetic products such as enzymes.
XX	
SQ	Sequence 19142 BP; 4629 A; 5539 C; 4716 G; 4238 T; 20 other:
	Query Match 19.2%; Score 33.2; DB 20; Length 19142;
	Best Local Similarity 53.0%; Pred. No. 0.5;
	Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
OY	9 agtgcgttagtgctaacgacactagaactctctgtcgtctctctccgcacagat 68
DB	3529 agagtcctcaaggctccctcaaatccgaactcgtcgtcctcttctggtccacgtacg 3588
OY	69 ctgcctctctcgtcgtcgtcgtacgaccctactcgtcgtcgtcgtcgtcgtcgtcgtc 128
DB	3589 ctgcgtgtcctcactcgtcgtctttagaacttttgtatctcgtcacagcatcgtcgtttg 3648
OY	129 ttcctctcctcaagt 142
DB	3649 ctgctcatgaact 3662
	RESULT 8
	N70435/c
XX	N70435 standard; cDNA; 1136 BP.
XX	
AC	N70435;
XX	
DT	05-APR-1991 (first entry)
XX	
DE	Sequence encoding insulin-like growth factor IB (IGF-IB).
XX	
KW	Growth promoter; lactation enhancer; cell proliferation; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP229750-A.
XX	
XX	22-JUL-1987.
XX	
XX	


```
PF 06-JAN-1987; 87EP-0870001.
XX
PR 20-NOV-1986; 86US-0929671.
PR 07-JAN-1986; 86US-0816662.
XX
PA (UNIW ) UNIV OF WASHINGTON.
XX
PI Kivli GG, Rotwein PS;
XX
DR WPI; 1987-200203/29.
XX
PT New pre-pro-insulin-like growth factor-1 protein - ohtd. by
PT recombinant DNA procedures for use as growth promoters for
PT enhancing lactation, for stimulating cell proliferation etc.
XX
PS Example; Fig 5; 59pp; English.
XX
CC A 42 base oligonucleotide corresponding to the DNA sequence encoding
CC amino acids 10 to 23 of mature human IGF-I was synthesized (N70437).
CC The radiolabeled 42 mer was then employed to screen for IGF-I
CC containing DNA sequences in a human liver cDNA library. Insulin-
CC like growth factors-1A and -1B cDNAs were isolated from a human cDNA
CC library by using lambdaagt 11 (N70435, N70436). The human IGF-1
CC genomic gene was isolated and mapped. It encodes at least two
CC preproinsulin-like growth factor-1 proteins. An essentially pure
CC preproinsulin-like growth factor-1 protein comprising the sequence
CC of amino acids shown in Figure six is claimed (P70277).
XX
SQ Sequence 1136 BP; 412 A; 230 C; 268 G; 226 T; 0 other;

Query Match 19.0%; Score 32.8; DB 8; Length 1136;
Best Local Similarity 52.1%; Pred. No. 0.3;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 34 agctctctgctcgtctctctctcgcgcacagatcgtctgctgctgctgctaga 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 727 ATCTCCCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 668

OY 94 accctaccctgctgcgagtgctgctgctgctgctgctgctgctgctgctgct 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 TTCTGCTTCCCTCTGAGTGTCTTTTGGCCAACTTCTCTCTCTCTCTCTCT 608

OY 154 gtgtgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 TTCTTGTGTAAGATGGGGGCTG 588

RESULT 9
A10594/c
ID A10594 standard; DNA: 10732 BP.
XX
AC A10594;
XX
DT 29-JUN-2000 (first entry)
XX
DE Gene encoding a subunit of cellulose synthase.
XX
KM Cellulose synthase; cellulose production; increase yield; ds.
XX
OS Vigna angularis.
XX
PN JP2000060568-A.
XX
PD 29-FEB-2000.
XX
PF 26-AUG-1998; 98JP-0239998.
XX
PR 26-AUG-1998; 98JP-0239998.
XX
PA (MIZU/) MIZUNO K.
PA (OUIP ) OUI PAPER CO.
XX
```

```
DR WPI; 2000-342371/30.
DR P-PSDB; Y85179.
XX
PT A gene encoding a cellulose synthetic equipment - for the improvement
PT in the amount of cellulose synthesised in a plant body
XX
PS Claim 2; Page 14-21; 32pp; Japanese.
XX
CC This sequence represents a gene encoding a subunit of the cellulose
CC synthase complex of Vigna angularis. The invention relates to subunits of
CC cellulose synthetic equipment, that can be used to increase the amount of
CC cellulose synthesised by a plant. The proteins and genes encoding them
CC can also be used to improve the properties of the cellulose being
CC produced by a plant.
XX
SQ Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

Query Match 18.2%; Score 31.4; DB 21; Length 10732;
Best Local Similarity 15.7%; Pred. No. 1.5;
Matches 27; Conservative 79; Mismatches 66; Indels 0; Gaps 0;

OY 1 gtaagccagatgctgctgctgctgctgctgctgctgctgctgctgctctcc 60
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
DB 9613 STTRCTBYSSRSRGYSYCSRSRRCYSCYTPDSYCYSTYAYSCYTSRGYSYDAS 9554

OY 61 gctcagatcgtctgctgctgctgctgctgctgctgctgctgctgctgctgct 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9553 RSTSTYSRCTTSTYSYSTTDYSDCSDYSTTBNSTYSSDSDBYSSDRCSRSDSTCNCY 9494

OY 121 gcttcctctctctcctcaagtcagatcgtgctgctgctgctgctgctgctgct 172
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
DB 9493 SCSDRSYSTTDACYTDAKTBCTYTDNCCNCSNSTYNTNSTYCSNTBSR 9442

RESULT 10
V21511/c
ID V21511 standard; DNA: 6755 BP.
XX
AC V21511;
XX
DT 17-AUG-1998 (first entry)
XX
DE Staphylococcal bacteriocin BacRI operon.
XX
KM BacRI; bacteriocin; antimicrobial; antibacterial; antibiotic;
KM Moraxella bovis; Infectious bovine keratoconjunctivitis; cancer;
KM therapy; ds.
XX
OS Staphylococcus aureus strain UT0007 (ATCC 55800).
XX
PN W09812319-A1.
XX
PD 26-MAR-1998.
XX
PF 18-SEP-1997; 97WO-US16758.
XX
PR 17-SEP-1997; 97US-0931999.
PR 19-SEP-1996; 96US-0710561.
XX
PA (UNIV ) UNIV KANSAS STATE RES FOUND.
XX
PI Crupper SS, Iandolo JT;
XX
PD WPI; 1998-230316/20.
XX
PT Therapeutic proteinaceous substances from Staphylococcus aureus -
PT useful to inhibit growth of wide range of prokaryotic or eukaryotic
PT cells, e.g. Moraxella bovis causing infectious bovine
PT keratoconjunctivitis
XX
PS Claim 2; Page 19-23; 38pp; English.
XX
```

CC This polynucleotide comprises the bacteriocin BacRI operon of
CC *Staphylococcus aureus* UT0007. The sequence of the BacRI operon
CC was determined by N-terminal sequencing of purified BacRI peptide
CC (see W54171), with back-translation and plasmid analysis. The
CC BacRI operon includes the BacRI gene (see V21510), a homologue of
CC the *cylM* gene of the cytolysin operon of *Enterococcus faecalis*
CC whose function is involved in the maturation of pre-cytolysin,
CC an ATP-transporter gene, *bio1* and *bio2* genes related to
CC lactococcal biosynthesis and modification, and a gene involved in
CC immunity function. BacRI peptides can be produced by construction
CC of an expression vector containing an oligonucleotide or operon
CC coding for BacRI, and use of the vector to transform host cells for
CC BacRI expression. The entire BacRI operon has been cloned into
CC plasmid pUB110, and *Bacillus subtilis* transformants secreted the
CC recombinant BacRI peptide into the medium. Bacteriocin BacRI is
CC active against many Gram-positive and Gram-negative organisms such
CC as *Bordetella bronchiseptica*, *Pasteurella multocida* and
CC *Staphylococcus aureus*; *Moraxella bovis*, causing infectious bovine
CC keratoconjunctivitis, is especially sensitive. BacRI can also be
CC used as an anti-cancer agent.

SQ Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U; 0 other;

Query Match 17.8%; Score 30.8; DB 19; Length 6755;
Best Local Similarity 54.4%; Pred. No. 2.1;
Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 36 ctctctgctcgcgtctctctcgcgcagatcgtctgctgctgctgctagaac 95

Db 4734 CTGTTTGGTGGTTTGTCCCTCTGCTTGTGCTTTTGTTCGTCCTCCCTCCCGT 4675

QY 96 cctactctgctgctgctgctgctgctgctgctcctcctcctcaagtcagtc 149

Db 4674 TCTGTGGTGTCTGTCTGTCTGCTGCTTCCCTCCCTCCCTTTCTTTT 4621

RESULT 11

ID X21686/c
X21686 standard; DNA; 1954 BP.

XX X21686;

DT 14-MAY-1999 (first entry)

DE WO9905284 Seq ID 32.

XX RecQ4 gene; helicase; Werner's syndrome; Bloom's syndrome; human; ss.

OS Homo sapiens.

XX Homo sapiens.

PN WO9905284-A1.

PD 04-FEB-1999.

PF 10-JUL-1998; 98WO-JP03114.

PR 25-JUL-1997; 97JP-0200387.

PA (AGEN-) AGENE RES INST CO LTD.

PI Furuchi Y, Kitao S, Shimamoto A;

DR WPI: 1999-142939/12.

PT New human helicase gene RecQ4 - used for investigation and diagnosis

XX of helicase-implicated diseases such as Werner's syndrome

PS Disclosure; Page 57-58; 67pp; Japanese.

XX The invention relates to a human gene RecQ4 encoding a protein having

CC helicase activity. The gene has significant homology to the *Escherichia*

CC coli helicase gene (RecQ). Host cells transformed with vectors comprising

CC the RecQ4 gene are used for the recombinant expression of the protein.
CC The gene may be used for the study and diagnosis of disorders in which
CC helicase activity is involved, such as Werner's and Bloom's syndromes in
CC which mutations in the helicase gene are implicated.

SQ Sequence 1954 BP; 369 A; 663 C; 607 G; 315 T; 0 other;

Query Match 17.2%; Score 29.8; DB 20; Length 1954;
Best Local Similarity 58.4%; Pred. No. 3;
Matches 52; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 55 ttctcgcctcagatcgtctgctgctgctgctgctgaacctactctgtcgcgagt 114

Db 1073 TTCTTCCGCGACATCTCTCCATGCTGCGGAGAGACCTGTACGAGTGCCTCG 1014

QY 115 gtgcgtgctgctgctcctcctcctcaagtc 143

Db 1013 CCCCACAGTAGTGTCTGCTTCATGTT 985

RESULT 12

ID X21656/c
X21656 standard; DNA; 3850 BP.

XX X21656;

DT 14-MAY-1999 (first entry)

DE Human helicase gene RecQ4.

XX RecQ4 gene; helicase; Werner's syndrome; Bloom's syndrome; human; ss.

OS Homo sapiens.

PN WO9905284-A1.

PD 04-FEB-1999.

PF 10-JUL-1998; 98WO-JP03114.

PR 25-JUL-1997; 97JP-0200387.

PA (AGEN-) AGENE RES INST CO LTD.

PI Furuchi Y, Kitao S, Shimamoto A;

DR WPI: 1999-142939/12.

DR P-PDB; W95050.

PT New human helicase gene RecQ4 - used for investigation and diagnosis

XX of helicase-implicated diseases such as Werner's syndrome

PS Claim 2; Page 35-42; 67pp; Japanese.

XX The present sequence represents a human gene RecQ4 encoding a protein

CC having helicase activity. The gene has significant homology to the

CC *Escherichia coli* helicase gene (RecQ). Host cells transformed with

CC vectors comprising the RecQ4 gene are used for the recombinant expression

CC of the protein. The gene may be used for the study and diagnosis of

CC disorders in which helicase activity is involved, such as Werner's and

CC Bloom's syndromes in which mutations in the helicase gene are implicated.

SQ Sequence 3850 BP; 711 A; 1250 C; 1273 G; 616 T; 0 other;

Query Match 17.2%; Score 29.8; DB 20; Length 3850;
Best Local Similarity 58.4%; Pred. No. 3.6;
Matches 52; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 55 ttctcgcctcagatcgtctgctgctgctgctgctgaacctactctgtcgcgagt 114

Db 1241 TTCTTCCGCGACATCTCTCCATGCTGCGGAGAGACCTGTACGAGTGCCTCG 1182

OY 115 gtgcgtctcgtctctcctcccaagtt 143
 Db 1181 CCCCACAGTAGTGTTCGTTCATGTT 1153

RESULT 13

X20167/c
 ID X20167 standard; DNA: 586 BP.

XX X20167;

DT 20-APR-1999 (first entry)

DE Enterococcus faecalis EF088 gene fragment.

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 detection; attenuation; antigenic; ss.

OS Enterococcus faecalis.

PN MO9850554-A2.

PD 12-NOV-1998.

PE 04-MAY-1998; 98WO-US08959.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;

DR WPI: 1999-070095/06.

DR P-PSDB; Y00177.

PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection

PS Claim 1; Page 181; 301pp; English.

CC The present sequence encodes an antigenic polypeptide fragment
 CC isolated from Enterococcus faecalis. The present invention describes
 CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
 CC The proteins can be used in vaccines for preventing or attenuating an
 CC infection caused by a member of the Enterococcus genus in an animal.
 CC They can also be used for detecting Enterococcus antibodies in a sample.
 CC The nucleotide sequences can be used for detecting Enterococcus nucleic
 CC acids. Products from the present invention can also be used for
 CC screening compounds to identify agonists and antagonists of E. faecalis
 CC protein activity.

XX Sequence 586 BP; 263 A; 106 C; 131 G; 86 T; 0 other;

Query Match 16.6%; Score 28.8; DB 20; Length 586;
 Best Local Similarity 50.7%; Pred. No. 4.4;

Matches 69; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 9 agtgcgttaagcctaagcgaacttaagcttctgcgtctgtctctctcgcgcacagat 68
 Db 179 ACTGCTGTTCTGCTTGGCTACTTCTGCTGCTGCTGCTGTTTGGATCGCCACAGTTT 120

OY 69 ctgcgttgcgttgcgtctgcgaacctactctgtgcgagtgctgcgtctgc 128
 Db 119 GTGCTTGTTCATGCTGCGTGTAGACTTTCACGCTGCTCGGTGTAGTCTTTTCTT 60

OY 129 ttctctctcaagtc 144
 Db 59 GGTGTTCTTTTGTTC 44

RESULT 14

X20162/c
 ID X20162 standard; DNA: 2559 BP.

XX X20162;

DT 20-APR-1999 (first entry)

DE Enterococcus faecalis gene EF086.

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 detection; attenuation; antigenic; ss.

OS Enterococcus faecalis.

PN MO9850554-A2.

PD 12-NOV-1998.

PE 04-MAY-1998; 98WO-US08959.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Bailey C, Choi GH, Hromocky J A, Kunsch CA;

DR WPI: 1999-070095/06.

DR P-PSDB; Y00172.

PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection

PS Claim 1; Page 176-177; 301pp; English.

CC The present sequence represents a gene isolated from
 CC Enterococcus faecalis. The present invention describes genes, proteins
 CC and antigenic polypeptides isolated from E. faecalis. The proteins can
 CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.

XX Sequence 2559 BP; 923 A; 427 C; 539 G; 668 T; 2 other;

Query Match 16.6%; Score 28.8; DB 20; Length 2559;
 Best Local Similarity 50.7%; Pred. No. 6.6;

Matches 69; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 9 agtgcgttaagcctaagcgaacttaagcttctgcgtctgtctctcgcgcacagat 68
 Db 2068 ACTGCTGTTCTGCTTGGCTACTTCTGCTGCTGCTGCTGTTTGGATCGCCACAGTTT 2009

OY 69 ctgcgttgcgttgcgtctgcgaacctactctgtgcgagtgctgcgtctgc 128
 Db 2008 GTGCTTGTTCATGCTGCGTGTAGACTTTCACGCTGCTCGGTGTAGTCTTTTCTT 1949

OY 129 ttctctctcaagtc 144
 Db 1948 GGTGTTCTTTTGTTC 1933

RESULT 15

X20164/c

ID		X20164 standard; DNA; 2559 BP.
XX	AC	X20164;
XX	DT	20-APR-1999 (first entry)
XX	DE	Enterococcus faecalis gene EF087.
XX	KW	Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
XX	KM	detection; attenuation; antigenic; ss.
XX	OS	Enterococcus faecalis.
PN		M09850554-A2.
XX	PD	12-NOV-1998.
XX	PF	04-MAY-1998; 98WO-US080959.
XX	PR	14-NOV-1997; 97US-0066009.
PR		06-MAY-1997; 97US-0044031.
PR		16-MAY-1997; 97US-0046635.
XX	PA	(HUMA-) HUMAN GENOME SCI INC.
PI		Balley C, Choi GH, Hiromocky} A, Kunsch CA;
XX	DR	WPI: 1999-070095/06.
DR		P-PSDB; Y00174.
PT		New isolated Enterococcus faecalis polynucleotides - used to develop
PT		products for the detection of Enterococcus and for use in vaccines
PT		for prevention or attenuation of Enterococcus infection
XX		Claim 1; Page 178-179; 301pp; English.
XX		The present sequence represents a gene isolated from
CC		Enterococcus faecalis. The present invention describes genes, proteins
CC		and antigenic polypeptides isolated from E. faecalis. The proteins can
CC		be used in vaccines for preventing or attenuating an infection caused
CC		by a member of the Enterococcus genus in an animal. They can also be
CC		used for detecting Enterococcus antibodies in a sample. The nucleotide
CC		sequences can be used for detecting Enterococcus nucleic acids.
CC		Products from the present invention can also be used for screening
CC		compounds to identify agonists and antagonists of E. faecalis protein
CC		activity.
SQ		Sequence 2559 BP; 923 A; 427 C; 539 G; 668 T; 2 other;
XX		
Query Match	16.6%;	Score 28.8; DB 20; Length 2559;
Best Local Similarity	50.7%;	Freq. NO. 6.6;
Matches 69; Conservative 0;	Mismatches 67;	Indels 0; Gaps 0;
QY	9	atgtgcttaagcgaacgaactagaagctcttgctgcgttcctctccgcgcagat 68
DG	2068	ACTCGTTTCCTCCTGGCTGCCTACTTCTGCTTGCGTTGCTGTTCATCACTGCCACAGTTT 2009
OY	69	ctgcttgcctgcctgcctgcctagaacctcactcgtgcctgcgaagtgcgtcgtcgtc 128
DG	2008	GTCGTTGTTGAATGCTTCGCGTGTGACTTTCACACTGCTGCTCGGTGTGAGTCTTTTCTT 1949
OY	129	tctctctctaagtc 144
DG	1948	GGTGTCTCTTTTGTGTC 1933

```
Search completed: February 12, 2001, 21:28:01
Job time: 3557 sec
```


Oy 61 gctcagaatcgcgtctgctgctgctgctgcgcagaaaccttaccgctgctgctggaatgctcgt 120
 |||||
 Db 61 gctcagaatcgcgtctgctgctgctgctgcgcagaaaccttaccgctgctgctggaatgctcgt 120
 |||||
 Oy 121 gctctgcctctcctcctcccaagttccgacatcgcgtatgctgctgctgaggggggcgcag 173
 |||||
 Db 121 gctctgcctctcctcctcccaagttccgacatcgcgtatgctgctgctgaggggggcgcag 173
 |||||

RESULT 2
US-08-750-007-1

Sequence 1 Application US/08750007
Patent No. 5801016
GENERAL INFORMATION:
APPLICANT: MORIOKA, SHINJI
APPLICANT: UKETI, JUN
TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT VECTOR
TITLE OF INVENTION: CONTAINING THE SAME AND METHOD FOR EXPRESSING FOREIGN
TITLE OF INVENTION: GENES USING THE SAME
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,007
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0760-221P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 183 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
OS-08-750-007-1

Query Match	100.0%	Score 173	DB 1	Length 183
Best Local Similarity	100.0%	Pred. No. 7.3e-48		
Matches 173	Conservative 0	Mismatches 0	Indels 0	Gaps 0

OY	61	gctcaagatcctgcattgttcgttgcctcgacgaaccctaactcgttgctcgaaatgtccgct	120
Dd	66	gctcagatcttcgcggctgctgccttcgcgatagaAACCCCTACTCgtctgcctgcGAGTgcGT	125
OY	121	gcttcgcgtctccctctcctccaagtctgcatactgctgctgtagtgaggggcgacag	173
Dd	126	gctttcgcgtcttccttcctccAAgcttccGAATCTAATTGTGTGTGGGGGCCAC	178

RESULT 3
US-08-446-794A-5

; Sequence 5, Application US/08446794A
; Patent No. 5747327

```

GENERAL INFORMATION:
APPLICANT: UERLI, JUNE
APPLICANT: MORIOKA, SHINJI
TITLE OF INVENTION: PHOSPHOLIPASE D GENE ORIGINATED FROM
TITLE OF INVENTION: PHOSPHOLIPASE D GENE ORIGINATED FROM
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,794A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0760-0203P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2799 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1876..1983
FEATURE:
NAME/KEY: CDS
LOCATION: 2524..2799
US-08-446-794A-5

```

Query Match	100.0%	Score 173;	DB 1;	Length 2799;
Best Local Similarity	100.0%	Pred. No. 2e+47;		
Match 173; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	61	gctcaagatcgcgtatgcgttgcgtcttcgcgcaaaccaactctgtgcgcgcagatgcct	120
Db	1726	gctcagatcgcgttcgcttcgcttcgcttcgcgcgaacccctactctgtcctccagatccct	1785
Qy	121	gcttcgctcttcctcctcaagatcgcatactgtagtgcgtgggggggcag	173
Db	1786	gcttcgcttccttcctcccaagttcgcattgcattgtgtgtgtgggggggcgcag	1838

RESULT 4
 US-08-750-007-4
 : Sequence 4, Application US/08750007
 : Patent No. 5801016
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: MORIOKA, SHINJI
 :
 : APPLICANT: UKI, JUN
 :
 : TITLE OF INVENTION: DNA FRAGMENT, RECOMBINANT VECTOR
 :
 : TITLE OF INVENTION: CONTAINING THE SAME AND METHOD FOR EXPRESSING FOREIGN

Query Match	15.8%;	Score 27.4;	DB 2;	Length 3531
Best Local Similarity	51.2%;	Pred. No. 5.8;		

1	NAME/KEY:	misc_feature	
2	LOCATION:	complement (937..2262)	
3	OTHER INFORMATION:	/label= ORF1SUBSEQUENCE	
4	FEATURE:		
5	NAME/KEY:	misc_feature	
6	LOCATION:	complement (2649..3458)	
7	OTHER INFORMATION:	/label= ORF3SUBSEQUENCE	
8	FEATURE:		
9	NAME/KEY:	misc_feature	
10	LOCATION:	3726..4799	
11	OTHER INFORMATION:	/label= ORF3SUBSEQUENCE	
12	FEATURE:		
13	NAME/KEY:	misc_feature	
14	LOCATION:	complement (4607..4918)	
15	OTHER INFORMATION:	/label= ORF5SUBSEQUENCE	
16	FEATURE:		
17	NAME/KEY:	misc_feature	
18	LOCATION:	5143..6216	
19	OTHER INFORMATION:	/label= ORF6SUBSEQUENCE	
20	FEATURE:		
21	NAME/KEY:	misc_feature	
22	LOCATION:	complement (5071..5643)	
23	OTHER INFORMATION:	/label= ORF7SUBSEQUENCE	
24	FEATURE:		
25	NAME/KEY:	misc_feature	
26	LOCATION:	6609..8888	
27	OTHER INFORMATION:	/label= ORF8SUBSEQUENCE	
28	FEATURE:		
29	NAME/KEY:	misc_feature	
30	LOCATION:	complement (6576..6830)	
31	OTHER INFORMATION:	/label= ORF9SUBSEQUENCE	
32	FEATURE:		
33	NAME/KEY:	misc_feature	
34	LOCATION:	9748..10044	
35	OTHER INFORMATION:	/label= ORF10SUBSEQUENCE	
36	FEATURE:		
37	NAME/KEY:	misc_feature	
38	LOCATION:	complement (10509..11282)	
39	OTHER INFORMATION:	/label= ORF11SUBSEQUENCE	
40	FEATURE:		
41	NAME/KEY:	misc_feature	
42	LOCATION:	12466..13002	
43	OTHER INFORMATION:	/label= ORF12SUBSEQUENCE	
44	FEATURE:		
45	NAME/KEY:	misc_feature	
46	LOCATION:	13723..14319	
47	OTHER INFORMATION:	/label= ORF13SUBSEQUENCE	
48	FEATURE:		
49	NAME/KEY:	misc_feature	
50	LOCATION:	15659..16210	
51	OTHER INFORMATION:	/label= ORF14SUBSEQUENCE	
52	FEATURE:		
53	NAME/KEY:	misc_feature	
54	LOCATION:	complement (16517..17545)	
55	OTHER INFORMATION:	/label= ORF15SUBSEQUENCE	
56	FEATURE:		
57	NAME/KEY:	misc_feature	
58	LOCATION:	complement (17737..18189)	
59	OTHER INFORMATION:	/label= ORF16SUBSEQUENCE	
60	FEATURE:		
61	NAME/KEY:	misc_feature	
62	LOCATION:	complement (18177..18743)	
63	OTHER INFORMATION:	/label= ORF17SUBSEQUENCE	
64	FEATURE:		
65	NAME/KEY:	misc_feature	
66	LOCATION:	complement (19031..19390)	
67	OTHER INFORMATION:	/label= ORF18SUBSEQUENCE	
68	US-08-008-216-19		

```

Query Match          15.3% Score 26.4; DB 1 Length 21126;
Best Local Similarity 52.8%; Pred. No. 23;
Matches 57; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY      44   tgcgtctgcttcctcccgcatcgtctgcttgcgtctgcgtacgaacctactct 103
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       3673 TTGCTTGCGACATCATTTTCTTAACCTGTGCCCTTTCTGTGGCCTCGCATCAATGCCGCCGA 3732
           ||||| || | | | | | | | | | | | | | | | | | | | | | |

QY      104   gtgcgcagagtgctgcgtctgcttccttcctccaagtgcattcga 151
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       3733 ACGAGCGACAGCCGTCGCCGATTGCCTTCCCAGAGGCGCAACTGGA 3780
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-459-569-19
Sequence 19, Application US/08459569
Patent No. 5543501
GENERAL INFORMATION:
APPLICANT: Slightom, Jerry L.
APPLICANT: Tepfer, David A.
TITLE OF INVENTION: R1 T-DNA Promoters
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: GRAY, CARV, AMES & FRYE
STREET: 401 B Street, Suite 1700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-4297

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,569
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,216
FILING DATE: 25-JAN-1993
APPLICATION NUMBER: US 06/725,368
FILING DATE: 22-APR-1985
ATTORNEY/AGENT INFORMATION:
NAME: Barthorst, Marnie W.
REGISTRATION NUMBER: 36,740
REFERENCE/DOCKET NUMBER: P1020U51
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 699-2700
TELEFAX: (619) 236-1048
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 21126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Agrobacterium rhizogenes
STRAIN: STRAIN A4
IMMEDIATE SOURCE:
LIBRARY: CONVOLVULUS ARVENSIS PLANT CELLS
CLONE: CLONE 7

FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (937..2262)
OTHER INFORMATION: /label= ORF5SUBSEQUENCE
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (2649..3458)
OTHER INFORMATION: /label= ORF2SUBSEQUENCE

```

```

1 FEATURE:
2 NAME/KEY: misc_feature
3 LOCATION: 3726..4799
4 OTHER INFORMATION: /label= ORF3SUBSEQUENCE
5 FEATURE:
6 NAME/KEY: misc_feature
7 LOCATION: 4401..4400
8 OTHER INFORMATION: /label= ORF4SUBSEQUENCE
9 FEATURE:
10 NAME/KEY: misc_feature
11 LOCATION: 4607..4918
12 OTHER INFORMATION: /label= ORF5SUBSEQUENCE
13 FEATURE:
14 NAME/KEY: misc_feature
15 LOCATION: 5143..6216
16 OTHER INFORMATION: /label= ORF6SUBSEQUENCE
17 FEATURE:
18 NAME/KEY: misc_feature
19 LOCATION: 5071..5643
20 OTHER INFORMATION: /label= ORF7SUBSEQUENCE
21 FEATURE:
22 NAME/KEY: misc_feature
23 LOCATION: 6609..8888
24 OTHER INFORMATION: /label= ORF8SUBSEQUENCE
25 FEATURE:
26 NAME/KEY: misc_feature
27 LOCATION: 6576..6830
28 OTHER INFORMATION: /label= ORF9SUBSEQUENCE
29 FEATURE:
30 NAME/KEY: misc_feature
31 LOCATION: 9748..10044
32 OTHER INFORMATION: /label= ORF10SUBSEQUENCE
33 FEATURE:
34 NAME/KEY: misc_feature
35 LOCATION: 10509..11282
36 OTHER INFORMATION: /label= ORF11SUBSEQUENCE
37 FEATURE:
38 NAME/KEY: misc_feature
39 LOCATION: 12466..13002
40 OTHER INFORMATION: /label= ORF12SUBSEQUENCE
41 FEATURE:
42 NAME/KEY: misc_feature
43 LOCATION: 13723..14319
44 OTHER INFORMATION: /label= ORF13SUBSEQUENCE
45 FEATURE:
46 NAME/KEY: misc_feature
47 LOCATION: 15659..16210
48 OTHER INFORMATION: /label= ORF14SUBSEQUENCE
49 FEATURE:
50 NAME/KEY: misc_feature
51 LOCATION: 16517..17545
52 OTHER INFORMATION: /label= ORF15SUBSEQUENCE
53 FEATURE:
54 NAME/KEY: misc_feature
55 LOCATION: 17737..18189
56 OTHER INFORMATION: /label= ORF16SUBSEQUENCE
57 FEATURE:
58 NAME/KEY: misc_feature
59 LOCATION: 18177..18743
60 OTHER INFORMATION: /label= ORF17SUBSEQUENCE
61 FEATURE:
62 NAME/KEY: misc_feature
63 LOCATION: 19031..19390
64 OTHER INFORMATION: /label= ORF18SUBSEQUENCE
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
```

[illegible]

Df Db

Qy 104 gtcgtcgagatgctgcgtcgttgcgtccctccctccaagtltgatctga 151
+ | |||| +||||| ++ | ||||| + || +
Db 3733 ACGACGGAGCATGCGCTGCCGGGATGTTCGCCTCCCGCAGGGCAACTGGA 3780

```

RESULT 14
US-08-458-831-19
: Sequence 19, Application US/08458831
Patent No. 5824866
GENERAL INFORMATION:
APPLICANT: Slightcom, Jerry L.
APPLICANT: Tepfer, David A.
TITLE OF INVENTION: Ri T-DNA Promoters
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: GRAY, CARY, AMES & FRYE
STREET: 401 B Street, Suite 1700
City: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-4297
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,831
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,216
FILING DATE: 25-JAN-1993
APPLICATION NUMBER: US 06/772,368
FILING DATE: 22-APR-1985
ATTORNEY/AGENT INFORMATION:
NAME: Barnhorts, Marlene W.
REGISTRATION NUMBER: 36,740
REFERENCE/DOCKET NUMBER: P1020US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 699-2700
TELEFAX: (619) 236-1048
INFORMATION FOR SEQ. ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 21126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Agrobacterium rhizogenes
STRAIN: STRAIN A4
IMMEDIATE SOURCE:
LIBRARY: CONVOLVULUS ARVENSIS PLANT CELLS
CLONE: CLONE 7
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (937..2262)
OTHER INFORMATION: /label= ORF1SUBSEQUENCE
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (2649..3458)
OTHER INFORMATION: /label= ORF2SUBSEQUENCE
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3726..4799
OTHER INFORMATION: /label= ORF3SUBSEQUENCE
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement (4041..4400)

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2001, 21:02:16 ; Search time 1923.53 Seconds
(without alignments)
630.245 Million cell updates/sec

Title: US-09-600-602-3

Perfect score: 173
Sequence: 1 gtaagccagtgctgcttagg.....gtgtgtgtggtggggcgccag 173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estro14:*
 121: em_estro15:*
 122: em_estro16:*
 123: em_estro17:*
 124: em_estro18:*
 125: em_estro19:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est67:*
 136: gb_est68:*
 137: gb_est69:*
 138: gb_est70:*
 139: gb_est71:*
 140: gb_est72:*
 141: gb_est73:*
 142: gb_est74:*
 143: gb_est75:*
 144: gb_est76:*
 145: gb_est77:*
 146: gb_est78:*
 147: gb_est79:*
 148: gb_est80:*
 149: gb_est81:*
 150: gb_est82:*
 151: gb_est83:*
 152: gb_est84:*
 153: gb_est85:*
 154: gb_est86:*
 155: gb_est87:*
 156: gb_est88:*
 157: gb_est89:*
 158: gb_est90:*
 159: gb_est91:*
 160: gb_est92:*
 161: gb_est93:*
 162: gb_est94:*
 163: gb_est95:*
 164: gb_est96:*
 165: gb_est97:*
 166: gb_est98:*
 167: gb_est99:*
 168: gb_est100:*
 169: gb_est101:*
 170: gb_est102:*
 171: gb_est103:*
 172: gb_est104:*
 173: gb_est105:*
 174: gb_est106:*
 175: gb_est107:*
 176: gb_est108:*
 177: gb_est109:*
 178: gb_est110:*
 179: gb_est111:*
 180: gb_est112:*
 181: gb_est113:*
 182: gb_est114:*
 183: gb_est115:*
 184: gb_est116:*
 185: gb_est117:*
 186: gb_est118:*
 187: gb_est119:*
 188: gb_est120:*
 189: gb_est121:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37.2	21.5	668	94	AM694929	AM694929 NF081E10S
2	37.2	21.5	671	94	AM694702	AM694702 NF079B05S
3	36.8	21.3	733	139	BF036458	BF036458 601460209
4	36.2	20.9	1078	190	CNS0074P	AL075925 DROSOPH11
5	35.8	20.7	974	137	BE880847	BE880847 601493284
6	35	20.2	743	109	BE519618	BE519618 HV_CED001
7	34.4	19.9	632	97	AM974965	AM974965 EST387070
8	34.2	19.8	990	193	CNS04YVF	AL313476 Tetradon
9	33.8	19.5	702	106	BE268416	BE268416 601124712
10	33.8	19.5	915	135	BE799838	BE799838 601584668
11	33.6	19.4	659	182	A2463907	A2463907 1M0273K03
12	33.6	19.4	997	190	CNS005TE	AI121302 uc30b03.r
13	33.4	19.3	443	16	AI121302	AI121302 uc30b03.r
14	33.4	19.3	584	27	AI981330	AI981330 pat. PK005
15	33.4	19.3	1265	139	BF032829	BF032829 601455658
16	32.8	19.0	473	10	AM703492	AM703492 z11h02.s
17	32.8	19.0	618	28	AJ398711	AJ398711 AJ398711
18	32.8	19.0	706	19	AI401719	AI401719 th30b10.x
19	32.8	19.0	821	176	A2202926	A2202926 SP_0079_B
20	32.8	19.0	876	192	CNS041KE	AL270311 Tetradon
21	32.8	19.0	893	105	BE259350	BE259350 601106438
22	32.8	19.0	900	105	BE249907	BE249907 600942852
23	32.6	18.8	693	160	AO656370	AO656370 Sheared D
24	32.6	18.7	746	107	BE415895	BE415895 MUG002.B0
25	32.4	18.7	903	94	AM728116	AM728116 GA_Ea001
26	32.4	18.7	1093	192	CNS03RTE	AL257676 Tetradon
27	32	18.5	484	92	AM577552	AM577552 PM2-B7054
28	32	18.5	876	175	A2138727	A2138727 SP_0170_A
29	32	18.5	998	135	BE799289	BE799289 601591751
30	32	18.5	1101	190	CNS017GS	AL108022 DROSOPH11
31	32	18.5	1201	190	CNS0167A	AL106384 DROSOPH11
32	32	18.5	1203	190	CNS015VR	AL106077 DROSOPH11
33	31.8	18.4	846	190	CNS010RJ	AL099337 DROSOPH11
34	31.8	18.4	897	138	BE973857	BE973857 601680665
35	31.6	18.3	356	193	FR0029482	AL025851 Fugu rubr
36	31.6	18.3	479	37	AV635060	AV635060 AV635060
37	31.6	18.3	620	180	A2358273	A2358273 1M0100E11
38	31.6	18.3	648	110	BE593348	BE593348 WSI_100_F
39	31.6	18.3	814	109	BE563541	BE563541 601334834
40	31.6	18.3	972	135	BE740157	BE740157 601595106
41	31.6	18.3	1235	137	BE881061	BE881061 601492122
42	31.6	18.3	1627	137	BE905523	BE905523 601495154
43	31.6	18.3	1822	138	BE967064	BE967064 601660603
44	31.4	18.2	353	146	W00065	W00065 T9EST74F0
45	31.4	18.2	360	39	AM086231	AM086231 xc70C10.x

ALIGNMENTS

RESULT 1
 LOCUS AM694929 668 bp mRNA
 DEFINITION NF081E10ST1F1082 Developing stem Medicago truncatula cDNA clone
 ACCESSION AM694929
 VERSION AM694929.1 GI:7569691
 KEYWORDS EST.
 SOURCE barrel medic.

ORGANISM	Medicago truncatula					
REFERENCE	Eukaryotes: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.					
AUTHORS	1 (bases 1 to 668) He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell ,C.J., Flores,H.R., Imman,J.T., Weller,J.W., May,G.D. and Dixon ,R.A.					
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation					
JOURNAL	Medicago truncatula stem library					
COMMENT	Unpublished (2000) Contact: Dixon RA Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7302 Fax: 580 221 7380 Email: radixon@noble.org Insert length: 668 Std Error: 0.00 Plate: 081 row: E column: 10 Seq primer: TCACACAGGAACAAGCATATGAC. Location/Qualifiers 1..668 /organism="Medicago truncatula" /db_xref="taxon:3880" /clone_id="NF081E10ST" /clone_lib="Developing stem" /lissue_type="stem" /dev_stage="Pooled developmental" /note="Vector: Lambda zap; Contains a mixture of internodal stem segments"					
BASE COUNT	330 a 62 c 188 g 88 t					
ORIGIN						
Query Match	21.5%; Score 37.2; DB 94; Length 668;					
Best Local Similarity	59.4%; Pred. No. 1.1;					
Matches	63; Conservative 0; Mismatches 43; Indels 0; Gaps 0;					
OY	48 ttctgtcttccgcgctcagcatctgtctgtctgtctgcgttcgaacactctgtgc 107 Db 432 TTCTCTGCCAATCGCTTCGGTCTGTGCTTTCTTCATCTCTTATCTTACTATTGT 373					
OY	108 tgcgagtgcgtcgcttcgtcttcctcccaagttcgatctgatt 153 Db 372 TTCACCTGTCTTCATCTCTTGCTTACTCTTTCCGCGTAATT 327					
RESULT 2						
LOCUS	AM694702/c 671 bp mRNA EST 15-JUN-2000					
DEFINITION	NF079B05SNI.F1044 Developing stem Medicago truncatula cDNA clone					
VERSION	NF079B05ST 5', mRNA sequence.					
KEYWORDS	AM694702 GI:7569464					
SOURCE	EST.					
ORGANISM	barrel medic. Medicago truncatula Eukaryotes: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.					
REFERENCE	1 (bases 1 to 671)					
AUTHORS	He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell ,C.J., Flores,H.R., Imman,J.T., Weller,J.W., May,G.D. and Dixon ,R.A.					
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation					
JOURNAL	Medicago truncatula stem library					
COMMENT	Unpublished (2000) Contact: Dixon RA Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7302					

	Fax: 580 221 7380
	Email: radixon@noble.org
	Insert Length: 671 Std Error: 0.00
	Plate: 079 Row: B Column: 05
	Seq primer: TCACACAGGAACAACACTATGTAC.
FEATURES	Location/Qualifiers
source	1..671
	/organism="Medicago truncatula"
	/db_xref="taxon:3880"
	/clone="NF07980557"
	/clone_1lb="Developing stem"
	/tissue_type="stem"
	/dev_stage="Pooled developmental"
	/note="Vector: Lambda zap; Contains a mixture of internodal stem segments"
BASE COUNT	314 a 80 c 180 g 97 t
ORIGIN	
Query Match	21.5%; Score 37.2; DB 94; Length 671;
Best Local Similarity	59.4%; Pred. No. 1.1;
Matches	63; Conservative 0; Mismatches 43; Indels 0; Gaps 0
QY	48 ttgtcttcctccgcacgatcgtcttgcttgctgctgctaaccactctgtgc 107
Db	272 TTCCTTCGCATCGCTTCCGTTGGTGCTTTTCATTCTCTTATCTTACATTTGT 213
QY	108 tcgcagttcgctgcgtctgctcttccttcctcaagtcatgatt 153
Db	212 TTTCACTTGTTTTCATTTCTCTTTCCTTACTCTTTTCCCTCGGTAATT 167
RESULT 3	
LOCUS	BF036458/c 733 bp mRNA EST 10-OCT-2000
DEFINITION	B01460209FI NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3863690 5', mRNA sequence.
ACCESSION	BF036458
VERSION	BF036458.1 GI:10744499
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 733)
AUTHORS	NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: DCTD/DMP
	cDNA library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
	Plate: LHCMS91 row: a column: 03
	High quality sequence stop: 685.
FEATURES	Location/Qualifiers
source	1..733
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:3863690"
	/clone_1lb="NIH_MGC_66"
	/tissue_type="adenocarcinoma"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."
BASE COUNT	175 a 195 c 208 g 155 t

[illegible]

FEATURES					
source					
9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 3528 Fax: 301 838 0208 Email: johng@ligr.org Plate: 343 Seq primer: Forward.					
location/Qualifiers					
1..632					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone_lib="MAGE resequences, MAGN"					
/note="Vector: pbluscriptSkm"					
BASE COUNT 88 a 163 c 105 g 276 t					
ORIGIN					
Query Match 19.9% Score 34.4; DB 97; Length 632; Best Local Similarity 52.9%; Pred. No. 6.9;					
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;					
OY	34	agctctgtcgcgcttcttcctccgcagatcagtctgtctgtctgtcgtcgtaga	93		
Db	447	ATTCCTCCCTGCTGCTGTCTTCTCCTGCATCTCGACACTGTGGTCCCTCC	506		
OY	94	acctactctgtgtcgtcgagtgctgcgtcttccttcctcaagttgatatt	153		
Db	507	TTCGTGTCCCTCCGTGATGCTCTTTGGCCAACTTCCTCTTGAGACTGCTGTT	566		
OY	154	gttgtgtgtggaggggcgacag	173		
Db	567	TTCGTGTGATGATGGGGGCTG	586		
RESULT 8					
CNS04YYE/c					
LOCUS					
DEFINITION					
CNS04YYF 990 bp DNA GSS 26-JUL-2000					
Tetraodon nigroviridis genome survey sequence T7 end of clone					
033E23 of library A from Tetraodon nigroviridis, genomic survey					
sequence.					
AL313476 GI:9546360					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
REFERENCE					
AUTHORS					
COMMENT					
Genoscope. Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon . Location/Qualifiers					

```
source
1. .990
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_1lb="053E23"
/clone_1lb="A"
/note="Genoscope sequence ID : COA053AC12c1-end : 17"

BASE COUNT      292 a      212 c      256 g      210 t      20 others
ORIGIN

Query Match
Best Local Similarity 51.7%; Score 34.2; DB 193; Length 990;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

OY 18 agagtaagcagactagagctctctgctgctgctctctctctccgcagatcgtcgt 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 574 AGTAGAGAGAGACCGCGGCTTTCTTCNTCCCTTCTTGTTATATCTCTTACT 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 78 tgcctgctcgtcagaaacctactctgctgtagagtgctgctcgtctcctcctcct 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 514 TGCCCTTTTCTTCCTCCCTCACACGCTCCTGGGCTGTGCTTAAATCTTCCTCC 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 138 caagtcgacatcgtatgtgtcgtg 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 454 TCTTCCCGCTGTGTTTGTCTTG 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
BE268416 702 bp mRNA EST 13-JUL-2000
LOCUS 601124712F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:2989466 5',
DEFINITION mRNA sequence.
ACCESSION BE268416
VERSION BE268416.1 GI:9142021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LHC480 row: g column: 03
High quality sequence start: 8
High quality sequence stop: 173.
Location/Qualifiers
1. .702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="IMAGE:2989466"
/clone_1lb="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT

328 a 142 c 190 g 42 t

```
Query Match
Best Local Similarity 56.9%; Score 33.8; DB 106; Length 702;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY 47 ctgtctctctccgcagatcgtctgtctgtctgtcgtcgtcgtcgtcgtcgtg 106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 CGTCCTTGTCTTCCTCGGCTCGATCTCTGTTTCCCTTCGTCCTCGGT 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 107 ctgagagtcgctgctgctgctcctcctcctcgaatcgcattgt 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 336 CGTCGCTTCGATTTGATGTTCTCTCCTCGTCAATCTCTTT 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
BE798838 915 bp mRNA EST 20-SEP-2000
LOCUS 601384668F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3939112 5',
DEFINITION mRNA sequence.
ACCESSION BE798838
VERSION BE798838.1 GI:10220036
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 915)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LHC787 row: g column: 17
High quality sequence start: 9
High quality sequence stop: 795.
Location/Qualifiers
1. .915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="IMAGE:3939112"
/clone_1lb="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT

282 a 200 c 276 g 157 t

OY 99

actctgctcgtcagatgctgctcctcctcctcgaatcgcattgtgtg 158

Db	738	AATCCATTCAAGTGTGAGCTCTTCCTCCTCTTTCTGGCTCCTTTCTTTTTCTTTT	679
QY	159	tgttgaggagcgc	171
Db	678	TTTGGGGCGCTTGC	666

RESULT	11
AZ463907	
LOCUS	659 bp DNA
DEFINITION	Muscle 10kb plasmid UUCGCM library Mus musculus genomic
ACCESSION	U00273K03 F. DNA sequence.
VERSION	AZ463907
KEYWORDS	AZ463907.1 GI:10622032 GSS.

REFERENCE
AUTHORS
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
1 (bases 1 to 659)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

TITLE	Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL	plasmid inserts
COMMENT	Unpublished (2000)
	Contact: Robert B. Weiss

84112, USA
 Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 Tel.: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0273 row: K column: 03
 Seq primer: CGTTGTAAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 659.

FEATURES	Location/Qualifiers
source	1. .659

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U081CM0273K03"
/clone_id="Mouse 10Kb plasmid U081CM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notice="Vector: pMD2.19. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson

```

laboratory mouse DNA resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The ligated DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g114732114[sh]AP123072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor-competent *E. coli* XL10-gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
ORIGIN

Query Match	19.48;	Score 33.6;	DB 182;	Length 659;
Best Local Similarity	63.88;	Pred. No. 12;		
Matches 51; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0;

```

Oy  7  ccagtgctgcttaggcgaagcgcaactagatgctctctgctgcttctcttcgcgccag 66
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 162 CAAGCATGGTGTAGGTGAGCGCATTTACAGCATCTGGCTTGCCTCAGGCTCTGCTGTGACCT 221

```

Qy	67	atctgcttgctgcttgc	86
Db	222	GTCCGCTTGCCTTGCCT	241

RESULT	12
CNS005TE	
LOCUS	CNS005TE 997 bp DNA GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR2K2 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION	AL060767	
VERSION	AL060767.1	GI:4943573
KEYWORDS	GSS.	
SOURCE	fruit fly.	

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)

REFERENCE AUTHORS TITLE JOURNAL	COMMENT
1 (bases 1 to 997) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)	Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp. the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	Location/Qualifiers
source	1. .997

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCL-98"
/clone="BACR12K23"
/note="end : TET3"
BASE COUNT      89 a      99 c      13 g      258 t      538 others
ORIGIN

```

Query Match	19.48;	Score 33.6;	DB 190;	Length 997;
Best Local Similarity	26.28;	Pred. No. 12;		
Matches 33;	Conservative 47;	Mismatches 46;	Indels 0;	Gaps 0

Oy 36 cttctgctgcgttccttctcctccgcgcagacatcgctctgcttgcttgcttagaac 95
::||| : :: |||::: | : | : : : :
... ..

D6 467 YTTCTGCTATYTTTATTATTTATTTTCATTTTCCACCCCAATTCCTTTTTTATTTATTTTYY 528
QY 96 cctactcgtgctgcagatgctgcgttcgtcttcccttcacagtctgatctgattgt 155

Db 527 CTCTGTTTCTTCATCAGTTTTTTTTTTTTTTTTTTCTCTCTTTTXX 586

QY 156 gtgtgt 161
: : |

JOURNAL
COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: DCRD/DTP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 plate: LICM579 row: d column: 16
 High quality sequence stop: 133.
 Location/Qualifiers

FEATURES
source

1..1265
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3859167"
 /clone_lid="NIH_MGC_66"
 /issue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

BASE COUNT 662 a 248 c 265 g 89 t 1 others
 ORIGIN

Query Match 19.3%; Score 33.4; DB 139; Length 1265;
 Best Local Similarity 53.4%; Pred. No. 15;

Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Oy 13 tgcctaggcctaagcgacactagagctctctgctgctctctctcgcgcacagatctgc 72
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1104 TgcTTGTCTGCTCTCGCGCTGTTCTTCTTCTGTTTCTTCTTCTTCTTCTGCTTCTGC 1045
 Oy 73 tgcctgctgctgcgacactagacactctgctgctgctgctgctgctgctgctc 132
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1044 GTTACTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTCG 985
 Oy 133 ttccctcaagct 143
 ||||| |||||
 Db 984 TTGCTCTGT 974

Search completed: February 12, 2001, 21:02:19
 Job time: 3975 sec

THIS PAGE BLANK (USPTO)